Assessment at the Crossroads: How Did We Get Here and What Should We Do?

Jane S. Halonen1, Melissa J. Beers2, and Anna N. Brown1

1 Department of Psychology, University of West Florida
2 Department of Psychology, Ohio State University

Psychology educators should have some advantages when it comes to responding to assessment planning and accountability mandates by virtue of their expertise in describing and measuring behavior. However, at this juncture, psychologists often show little enthusiasm for embracing assessment as part of their obligation in the teaching role. In this article, we offer some history regarding how assessment became such a dominating but unpopular force in higher education. We explore the lost promise of assessment from multiple perspectives, including the public, accrediting agents, administrators, faculty, and students. We argue for steps psychology programs can take to develop a true assessment culture and meaningful improvement.

Keywords: college teaching, assessment, higher administration

In principle, no psychologist should object to the assessment of student learning. Our science is predicated on the assumption that measurement helps us determine the impact of relevant variables on behavior. It is paradoxical that assessment of learning in higher education—especially in psychology courses—generates so little enthusiasm, especially among experts in measuring, describing, explaining, and influencing behavior and mental processes.

Ikenberry and Kuh (2015) described how assessment practices became a standard feature in higher education and bemoaned the fact that colleges and universities were having difficulties establishing cultures where assessment firmly takes root. In a true culture of assessment, all stakeholders should be enthusiastic about the opportunity to generate evidence of student learning. Yet contemporary assessment practices do not reflect this positive perspective. Instead of being a fruitful process that improves inclusion and student academic achievement, most stakeholders view assessment as fatiguing, uninformative, and ineffective—and justifiably so. Although the assessment-reform movement may have had good intentions from the outset, critics today emphasize that higher education’s accountability focus misdirects attention, encourages unethical behavior, wastes time and money, and creates tensions across stakeholders.

Our purpose in writing this article on the general status of assessment is twofold. First, we want to promote the position that assessment, done for the right reasons, represents an essential tool that benefits all stakeholders in higher education, including the public, accrediting bodies, administration, faculty, and students. We will examine the perspectives of these stakeholders in their efforts to launch, sustain, support, or participate in assessment culture and conclude by offering some ideas on how to build and nourish helpful assessment practices in psychology programs. Second, we hope our overview will provide a fitting back-
drop for the articles that constitute this special issue of *Scholarship of Teaching and Learning in Psychology*. The articles that follow present assessment exemplars done for the right reasons. But first, we explore the reasons that assessment often misses the mark.

**The Public Demands More of Higher Education**

Historians of higher education point to the impact of a presidential initiative that led to a report called *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education, 1983) as a tipping point for serious public concern with quality in higher education (Kamenetz, 2018). *A Nation at Risk* was not the first presidential action to address educational shortcomings in America. All mid-20th-century presidents appointed commissions to ensure that education was serving American interests effectively (e.g., Truman’s Commission on Higher Education [1947]; Eisenhower’s Committee on Education in High School [1956]; Kennedy’s Task Force on Education [1960]). However, Ronald Reagan’s *A Nation at Risk* sounded an alarm that American schools had failed in their shared mission to equip citizens with the tools they would need to help America prevail in a competitive global environment, and Americans took notice.

*A Nation at Risk* catalogued a long list of criticisms to inspire corrective action, including blame to share across all levels of education. Regarding higher education, commissioners concluded that declining SAT scores represented “a rising tide of mediocrity that threatens our very future as a nation and as a people” (National Commission on Excellence in Education, 1983, p. 1). *A Nation at Risk* recommended 38 specific actions to improve conditions, including raising standards for admission, guarding against grade inflation, using standardized tests as a benchmarking strategy, and increasing federal oversight to monitor changes and promote improvement.

Critics of the report identified many problems with *A Nation at Risk*, including the distorting effects of confirmation bias. Kamenetz (2018) interviewed the original commissioners, who confessed that they intentionally used alarmist language to make a persuasive case for action. America was experiencing a recession, and education received paltry attention in contrast with larger concerns with the economy. Commissioners wanted to rally Americans to care more about producing a better-prepared workforce as one solution to our economic doldrums.

Guthrie and Springer (2004) also claimed the statistical analysis of *A Nation at Risk* was flawed from “cooking the books” to produce the outcome the commissioners wanted. For example, the commissioners offered declining average SAT scores as proof of deteriorating quality, but they failed to take into account that broader successful tactics to recruit underserved populations expanded the range of SAT scores in a negative direction, exerting a downward skew on the average score. Each racial subgroup had actually demonstrated improvement, but taken collectively, the data produced the illusion that the average score was declining. Regardless of the potential statistical sleight of hand, the alarmist language predicting looming disaster resonated with the public.

Public stakeholders have shown increasing concern regarding whether the cost of college is justified because costs in higher education routinely outstrip inflation. According to Maldonado (2018), student debt now represents the largest proportion of American debt after mortgages. He calculated the cost of a 4-year degree across different types of institutions at an average of over $104,000. Burgeoning student debt is even less acceptable to the public if the end product, the college-degree-holding student, cannot meet the basic workforce needs in problem solving, critical thinking, and communication.

**Accountability Mandates and Assessment**

One corrective action recommended by *A Nation at Risk* that flourished was empowering specific organizations to provide oversight for accountability concerns. Higher educators must now contend with three different groups that address accountability: regional accrediting bodies, state regulations, and professional organizations. Accountability agents tend to focus on quality assurance and institutional improvement, but each stakeholder may approach these issues from such different perspectives that educators can experience conflict in expectations across their respective assessment overlords.
Regional Accreditation Processes

Although the first regional accrediting body was established in 1885, it was not until the 1980s that Secretary of Education William Bennett required accrediting bodies to review explicit learning gains as part of the process of securing regional approval (Ewell & Jankowski, 2015). Prior to that mandate, accrediting bodies focused on resources and regulations or “inputs” (e.g., volumes in the library, credentials of the faculty, incoming student board scores) rather than “outputs” (i.e., what students actually achieved during their collegiate experiences).

According to Ewell and Jankowski (2015), the Southern Association of Colleges (SACS) instituted the first requirements for its members to offer evidence of student learning, an initiative that was quickly adopted by all other regional bodies. Although the regional accreditors did not stipulate a specific standard for accountability practices, by 1990, all accreditors expected their institutional members to develop an accountability strategy—an assessment plan—to examine what students actually learn in their respective programs.

In the early days of accreditor reviews, institutions merely needed to outline a plan for collecting assessment data to satisfy accreditor oversight (Halonen, 2018). However, requirements evolved over time to include the collection of outcome data and the use of those data to “close the loop” to foster data-based decisions as drivers in decisions about program improvements. Halonen (2018) also described that accreditor demands have become more nuanced over time. For example, programs may need to disaggregate performance data according to whether students are “first time in college” or transfers and whether classes are online or face to face. Accreditors may request performance data on certificate programs, general education offerings, and minors—a focus that goes far beyond the original mission of accreditation oversight.

Why have accrediting agencies become so powerful? Federal regulations (e.g., see https://www.ed.gov/category/keyword/accreditation) make student access to federal sources of funding contingent on educational programs’ maintenance of regional accreditation status. Consequently, educational institutions must go through a formal review process, typically once a decade, providing assurance that institutions expend federal funds in appropriate ways. Universities now submit midpoint reports to reduce the predictable flurry of assessment activity preceding an accreditation site visit in an attempt to avoid long stretches of time when accountability practices tend to recede in importance and, in many cases, wither.

What started out as fairly simple expectations for assessment evidence (i.e., What are your students learning, and what can they do as a result?) has mushroomed into an extravagant enterprise that occupies a great deal of time and attention. The accrediting process has generated significant criticism as being overcomplicated, nontransparent, expensive, ineffectual, and driven by standards geared to the minimally capable (Banta & Palomba, 2014).

Legislative Oversight

For some institutions, other accountability stakeholders capable of issuing influential mandates are state legislatures and boards of education appointed to govern educational quality in state-subsidized programs. For publicly funded institutions, the magnitude of funding may be contingent on institutional compliance with what the state requires. Legislative reach may entail specifying outcome requirements and general education options, among other mandates.

Little uniformity exists across the states in how educational programs should perform. However, many legislatures have adopted some version of “performance metrics” as a strategy to promote high-quality programs. For example, universities may need to report first-year retention rates, graduation rates, and even the average salary of graduates 6 months after graduation. Institutions that do not meet specified standards may be punished with budget reductions, whereas those achieving specified outcomes may be granted additional funding to reward their efforts—or not. Rewards for effective performance have not been reliable. These strategies have transformed assessment into a high-stakes activity.

Muller (2019) suggested that stakeholders may respond to metric demands by gaming the system. For example, admissions officials and administrators have collaborated to lower stan-
standards to improve enrollment capture or to increase standards to improve completion rates. Faculty may experience similar pressures from metric tyranny to be less rigorous in grading practices to enhance student evaluations of teaching. None of these efforts achieves the goal of improving learning outcomes.

**Professional Standards in Psychology**

Disciplines should know their needs best, and many have generated professional standards for outcomes that should influence curricula. According to Ewell and Jankowski (2015), approximately 50 formal accreditors certify quality in such diverse fields as chemistry, engineering, medicine, and psychology. These organizations may set forth a preferred curriculum, specify relevant learning outcomes, and dictate specific types of performance data, such as a comprehensive exit exam. The emphasis of these bodies tends to focus on maintaining program quality rather than necessarily promoting continuous improvement objectives.

Although psychology does have an accreditation process for clinical programs at the doctoral level, the American Psychological Association (APA) has not been eager to embark on such an expensive and time-consuming process at the undergraduate level. However, the APA’s Board of Educational Affairs initiated many activities to support faculty facing accountability demands, including task force assignments, policy statements, and conferences design to enhance the quality of the educational experience in psychology. Table 1 summarizes the initiatives sponsored by the APA in the last 2 decades designed to enhance high-quality curriculum and assessment practices. Two more initiatives are under way. The APA Introductory Psychology Initiative will introduce a national agenda about the design and delivery of the first course in the curriculum. As well, discussions are underway for Guidelines 3.0. Both projects should come to fruition in 2021.

**Administrative Dilemmas in Assessment**

Administrators are understandably challenged in providing evidence of learning that satisfies diverse audiences as well as in determining how to communicate and support the rationale for these requirements to the faculty. In the accountability food chain, administrators are caught in the middle. According to Ewell and Ikenberry (2015), academic leaders must oversee the process that gathers evidence of learning, although they are typically several steps removed from data gathering and interpretation. They may be untrained in assessment and perhaps even philosophically opposed to the process. However, their responsibilities require accepting assessment initiatives from above and helping those on the front lines of teaching and learning get on board. Ewell and Ikenberry claimed that responsiveness to assessment might distinguish whether administrators are serving as leaders or merely managers of the educational climate.

For this stakeholder discussion, we broadly define “administrators” as those with supervisory or leadership responsibilities that influence how assessment practices will unfold for their constituents. Program chairs and directors, deans, provosts, and presidents all have a stake in how well their programs fare under scrutiny. Assessment demands have burgeoned into an enterprise that requires substantial time, energy, and money, complicated by a variety of administrative challenges.

**The Unfunded Mandate Issue**

Assessment began as a simple request for data that was relatively easily accommodated. However, as data demands evolved, complex financial decisions followed. Staff positions established to oversee and manage the assessment process may require reallocation of funding for new lines. Learning-mediated systems to collect required data add expense and complexity to the process. Adopting a nationally normed exit exam will require funding; sometimes that cost is passed on to the student in terms of a lab or student fee. Although data demands continue to grow, rarely do new streams of financial support materialize to address the costs of meeting those demands.

A related financial problem administrators face is resource allocation. Accredited programs are in a better position to ask for scarce resources with the attendant threat of “losing accreditation” if their needs go unmet. Disciplinary standards can specify requirements that include the square footage required to deliver a program, desirable faculty-to-student ratios, and
### Table 1

**Contributions of the American Psychological Association (APA) to Undergraduate Curriculum and Assessment in Psychology**

<table>
<thead>
<tr>
<th>Year</th>
<th>Resource</th>
<th>Purpose</th>
<th>Relevant links</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003</td>
<td>Guidelines for the Psychology Undergraduate Major</td>
<td>Identified first national guidelines on student learning outcomes in psychology</td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>Teaching, Learning, and Assessing in a Developmentally Coherent Curriculum</td>
<td>Identified appropriate outcomes and strategies for community college classes</td>
<td><a href="https://www.apa.org/ed/governance/bea/undergrad-curriculum">https://www.apa.org/ed/governance/bea/undergrad-curriculum</a></td>
</tr>
<tr>
<td>2010</td>
<td>Undergraduate Education in Psychology: A Blueprint for the Future of the Discipline</td>
<td>Book drawn from proceedings of the 2008 conference</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>Principles for Quality Undergraduate Education in Psychology</td>
<td>Principles extracted from Blueprint document</td>
<td></td>
</tr>
</tbody>
</table>
specific training backgrounds for program administrators hired to run accredited programs. Faced with a conflict between two equally needy programs, the program that risks losing accreditation is likely to win. Because the APA has been resolute about not pursuing accreditation at the undergraduate level, psychology department chairs simply do not have the same leverage as accredited programs to fund assessment efforts for the sake of educational improvement.

The Mission-Creep Problem

As accrediting demands have matured, the scope of data collection has dramatically expanded. For example, regional accreditors initially wanted institutions to provide self-generated evidence that the university under review was living up to its mission statement. Without imposing specific data-collection requirements, accreditors focused on the value of an undergraduate degree. However, universities discovered that each accrediting visit ushers in new emphases in reporting requirements, requiring more data collection and interpretation—for example, “Are general education classes adding value to the degree?” Each new question imposes more demand on administrators to energize faculty to deliver broader evidence that the program is succeeding in meeting all of its institutional claims.

In large departments, mission creep means that data collection may have to transpire in multiple contexts. Faculty working in universities that run remote campuses or online programs may need to provide evidence that those experiences are equivalent to face-to-face programming on the main campus. Departments discover that each accrediting visit ushers in new emphases in reporting requirements, requiring more data collection and interpretation—for example, “Are general education classes adding value to the degree?” Each new question imposes more demand on administrators to energize faculty to deliver broader evidence that the program is succeeding in meeting all of its institutional claims.

The Mixed-Motive Problem

Two distinctive purposes can be served by the process, distinguished as the difference between assessment for learning versus assessment of learning (McCarthy, 2015). In assessment for learning, assessment results are returned to students so that they can benefit and grow from the feedback inspired by their achievement. In contrast, many assessment activities increase the level of analysis to program evaluation or assessment of learning. Faculty gather and interpret data from courses to document how the course or program as a whole is performing. Consequently, feedback is not necessarily returned to the student. In such cases, administrators face the challenge of optimizing student engagement when there may be no direct benefit to the student for participating in those assessment activities.

The Oversight Challenge

Management of the assessment process may be assumed by a department chair; however, it is more likely that another individual will be drafted to serve in a coordinating capacity. As the tasks expand, coordinators may appropriately ask for that burden to be reflected in workload, either through extra compensation or release time. However, budgets are rarely carved out at the department level to accommodate additional continuous expense for responsibilities that become more burdensome over time. Even when departments are well prepared to address assessment concerns, the expansion of requirements may entail placing extra data-collection burdens on full-time and part-time instructors or staff members.

In many cases, the dramatic growth of assessment demands required instituting new administrative positions to oversee the process, accounting for a significant portion of what higher education critics complain about as “administrative bloat” (Williamson et al., 2018, p. 15). Assessment officials and their staffs can make demands on instructors’ time that are independent of the direction they receive from their immediate supervisors. For instance, they may call meetings to discuss and review assessment results across programs or issue new administrative forms to complete as part of an annual process.

The Complexity of Harvesting Results

Form should follow function, but reporting-form design may take place with little to no input from those who will ultimately implement the process. To satisfy outside audiences and make compilation efficient, reporting forms tend to be standardized and quantified. Conse-
quently, data-collection procedures look for common denominators. Ironically, adopting a metrics-oriented framework may discourage intrinsic motivation to pursue improvement goals. Assessment leaders may impose strict reporting processes that might not necessarily fit with how discipline managers would prefer to tell their stories of what they have been achieving.

Gathering data efficiently poses some unique challenges. Recent features of learning-mediating systems have recognized that automating the gathering of data on outcomes would be a useful feature for faculty. The infrastructure of those systems may impose limitations on how and what can be assessed. Should faculty wish to pursue an assessment question that does not fit well with the existing capabilities of the system, their research questions about student achievement of outcomes will be stifled or may require costly upgrades to the basic features of the learning-management system.

The Generation of Meaningful Results

The early faculty response to assessment needs focused on grade distributions as evidence of learning quality. However, critics of that approach point out that a grade distribution does not provide sufficient information to “close the loop” (Lynch & Hennessy, 2017). How would observing the number of Bs versus Ds in a class help an instructor improve the course experience? A shift toward specific behavioral outcomes resulted. In this regard, psychologists should technically have a clear advantage in understanding how to articulate and promote expected behaviors resulting from classes, yet as Dunn et al. (2020) point out in this issue (pp. 244–253), psychology fares no better in advancing the utility of assessment in its courses than any other discipline.

High-quality education is challenging to assess, so numbers that are easy to capture (e.g., retention of first-year students, average salary 6 months after graduation) dominate data-collection strategies and may substitute for the harder work of defining and measuring real academic achievement. Fortunately, quantification requirements usually play to our discipline’s strengths.

Defining acceptable goals is also not easily accomplished. Should programs be satisfied if 75% of students achieve a desirable outcome? Or should we be satisfied if students can achieve 75% of the expectations that have been articulated? Is the desirable outcome predicated on what the average student should know and do? We can establish improvement goals that monitor the same goal over years to see the trends over time or change the emphasis once a goal is achieved. Accountability practices embody the spirit of continuous improvement; consequently, they require faculty to be on a constant problem- and solution-finding mission. A high-functioning program could be justified in establishing and maintaining high-level achievement from year to year to satisfy accountability demands, but an emphasis on continuous improvement may simply feel like an endless game in which the goalposts are continuously receding.

The Feedback Burden

Annual reporting practices typically require concise data summaries that must be prepared for audiences with varying backgrounds in statistical analysis. Therefore, psychology administrators must be especially careful to translate their interpretations of the patterns observed in assessment data into language that can be understood by those who do not know and are less likely to care about the data challenges of oversized standard deviations or undersized samples. Psychologists often have an advantage over colleagues in other disciplines where quantitative metrics are concerned, but we also may overlook the value of qualitative measures and content analysis to inform assessment efforts.

Creating reports for stakeholders is a time-consuming process that is often uncompensated and unrecognized. Many who toil to produce assessment reports believe the reports are unlikely to be read or understood. An English professor colleague of the senior author disclosed that in one annual report’s discussion of performance metrics, he had included a chart depicting tractor production. No one noticed. Such episodes appropriately fuel cynicism on the part of those gathering and interpreting the data that they are participating in a hollow exercise.

The Morale Morass and Failed Promises

Faculty and other school officials suffer “initiative fatigue” from the varying, unpredictable, and seemingly relentless institutional agenda
items that continuously emerge (Borray & Millichap, 2017). Without predictable reinforcement for effort and with the constant distraction of the next new and shiny goal, enthusiasm for accountability practices will dwindle. Wise administrators enact assessment plans as leaders, rather than merely managers, to keep faculty engaged and inspired.

Administrators advocate that producing strong metrics generates the best argument for maintaining or expanding resources. However, sudden budget shortfalls or emerging new priorities can divert original plans and promises. In addition, data that do not support preexisting administrative biases about program quality can be ignored if administrators give greater weight to the accuracy of their own intuitions about program quality over whatever the data reveal. An additional problem is that when performance metrics divide up available funding according to the quality of the data, it promotes competitive rather than collaborative relationships among programs.

A Final Administrative Irony

Administrators rarely face the same kind of evaluative scrutiny that teachers and students do in the assessment process. However, the success of an institutional assessment plan can be construed as a measure of how well they perform in a leadership capacity. Completing a positive assessment process and submitting a report on time do not predictably translate into new resources, but administrators can find themselves under threat if their programs do not measure up or if their reports are chronically delayed or missing. Avoiding punishment is unlikely to be a strong incentive for administrators to become assessment enthusiasts. In the following section, we offer additional considerations to broaden these recommendations and support teachers of psychology at all levels.

Teacher Assessment Challenges

One might expect psychology teachers to be at least intrigued by assessment, if not its best advocates, but as Chew et al. (2018) point out, we do not always practice what we teach. Dunn et al. (2020) deftly mapped out an array of factors contributing to misunderstandings and missed opportunities for teachers of psychology to leverage assessment toward improved student learning. The “tyranny of content” that plagues many psychology courses, ever-increasing enrollments, the prevalence of multiple-choice testing, a historical lack of training and support for how to align assessments with learning outcomes and skills, a lack of recognition for teaching and assessment as scholarship, and a dearth of support and encouragement for the enterprise all provide obstacles to enthusiasm for assessment. These conditions, mixed with external pressures from administrators and accreditors, have created a fallow field for assessment in higher education when assessment should be a natural part of the teaching and learning process.

The Contingency Crisis

Broadening the view from the landscape painted by Dunn et al. (2020), we see several additional elements in teachers’ higher education ecosystems that stifle the potential for growth of true assessment culture. First, on a national level, adjunct and teaching-track faculty comprise an ever-increasing proportion of undergraduate teaching. These faculty positions are referred to as “contingent” because they offer little predictability, protection, or security. According to the American Association of University Professors (AAUP, 2006), in the United States, 73% of instructional positions are contingent.

Contingent faculty face special challenges in higher education. Teaching faculty and graduate student instructors can be hired weeks or days before the start of a class, with little if any time or incentive to develop courses that dovetail with the requirements of the program or institution. These faculty positions are referred to as “contingent” because they offer little predictability, protection, or security. According to the American Association of University Professors (AAUP, 2006), in the United States, 73% of instructional positions are contingent.

Contingent faculty face special challenges in higher education. Teaching faculty and graduate student instructors can be hired weeks or days before the start of a class, with little if any time or incentive to develop courses that dovetail with the requirements of the program or institution. Contingent faculty often do not know when they might teach a particular class again. Contingent faculty are typically not included in curricular planning or decision making alongside tenure-stream colleagues. Higher education treats contingent teaching professionals as interchangeable, which undermines their ability to contribute to assessment.

The Impact of Negative Motivation

The contingency problem demotivates teachers. At many institutions, particularly large, research-intensive universities, faculty advancement is based on research productivity
and graduate training, leaving little time and few resources for teaching. Consequently, to protect the time of tenure-track faculty, administrators frequently delegate teaching and assessment activities to professional or administrative staff, contingent faculty, and even graduate students. These roles are subject to high rates of turnover. A common practice involves installing an individual from these groups in the role of an “assessment coordinator.” Allocating the work of assessment to someone other than faculty, especially if those individuals represent positions with less status and prestige, may be construed as absolving faculty of the responsibility to assess their students’ learning. In the eyes of some faculty, consigning assessment tasks to nontenured roles discredits the assessment process as invalid or unimportant.

**The Empty Allure of Academic Freedom**

Balancing individual instructors’ areas of expertise and intrinsic interest/motivation with consistency in assessment adds layers of complexity to an already-challenging task. Particularly for core courses in the major, such as introductory psychology, research methods, data analysis, and courses in the institution’s general education curriculum, different individuals may teach multiple sections of the same courses. They may teach the courses as completely independent sections with no coordination at all. Alternatively, they may be required to teach a tightly standardized course designed to ensure consistency across sections. Coordinated efforts to set educational priorities strengthen and streamline assessment (Beers, 2020); however, this practice flirts with threats to academic freedom, according to critics who oppose time invested in course alignment.

The concept of academic freedom centers on protecting the rights of teachers in how to approach their subject (AAUP, 2006). Yet, academic freedom does not imply that individuals should conduct their teaching or their academic work alone. In teaching, just as in research, an integrated, collaborative, and inclusive effort to develop and enact a curriculum should produce stronger and more robust results for student learning in the long term.

In a context of mutual respect and shared decision making, no one individual “owns” a course entirely. Each course makes both a unique and shared contribution to students’ overall experience in a program. This barrier to assessment is ultimately less about understanding and implementing assessment than about enacting sustained collaborative teaching practices. Teachers of psychology likely recognize that science is stronger when scholars collaborate (Liao, 2011), so why do not we collaborate when it comes to our teaching? Why do we do so much teaching with our doors closed?

**Shifting Our Culture to Open Teaching**

In recent years, frustration with unreplicable results from published reports and a desire to strengthen the accuracy, reliability, and veracity of scientific findings led to a movement toward “open science” (van der Zee & Reich, 2018). Open science aims to increase the transparency of and access to research findings. Researchers openly share their hypotheses, research designs, analysis plans, and data for dialogue and iterative feedback from the wider scientific community. Teaching stands to benefit tremendously from the lessons learned in the open-science movement, but for some reason, psychological scientists seem reluctant to apply the same practices they hold near and dear in their research pursuits to their teaching activities. Scientists agree that we should not start collecting data without a sound and well-reasoned hypothesis, but scientists as teachers may not always recognize the value of backward design to orient and align their courses and assessments with objectives for student learning.

Teachers are likely to collect assessment data post hoc and without a priori integration in course and curricular planning. Opportunistically drawing conclusions about student learning at the end of a course, in the absence of sound design, smacks of HARKing, or hypothesizing after the results are known, which inflates the risk of Type I error (Kerr, 1998). Scientists avoid conducting studies haphazardly, without hypotheses, and drawing conclusions after the fact, but we suspect that a fair number of scientists teach courses without systematic planning and assessment. Scientists expect evidence to undergo peer review but are less likely to ask for input on the design of assessments or their proper alignment to the
course goals. Teaching courses in isolation from others in the curriculum and without assessment is like “throwing darts with your eyes closed” (Landrum, 2016). Even if you generate data from such an exercise, the benefits for student learning in the program are uncertain. Indeed, students are the ones who pay the price for lackluster assessment efforts.

**Students Under Assessment Scrutiny**

Despite the fact they are the focus of assessment efforts, consideration of student perspectives and experiences routinely seems to get lost in the shuffle. In this final stakeholder section, we address why assessment may fail to support improved learning from the student point of view.

**The Problem of Misguided Motives**

A fundamental issue regarding the success of student assessment practices involves how students view learning. Many students—perhaps the majority—simply approach learning in a different way than educators do. Unfortunately, students get caught up in the quest for a certain grade, whether it is a shiny A or a serviceable C. Many feel compelled to learn only the minimum content necessary to do well on a test or other assignment, engaging in surface learning (cramming and memorization) or strategic learning (superficial study methods designed to maximize points; Struyven et al., 2005). In the short-term press of preparing to be graded, they forget the nobler, long-term objectives of becoming better thinkers, writers, and speakers, limiting their ability to make deep connections to the course content so that long-term learning can take place. Teachers may not especially emphasize the value of what the course has to offer beyond the end of the term, leaving little motivation to attempt deep learning and reinforcing students’ propensity to care more about grades than learning.

Undergraduate programs in psychology may be particularly vulnerable to this problem for multiple reasons. Professors’ inherent enthusiasm for the specialized area in which they teach may encourage several kinds of disconnection. Each individual class session may be disconnected from other sessions in the course; the course may be disconnected from other classes that comprise the student’s major; and perhaps most critically when concerns exist about whether psychology majors facilitate securing jobs, the course may be disconnected from the demands of potential careers after graduation.

Early in students’ education, much of their class time is spent on standardized testing, shaping how students view learning (Strauss, 2015). From the perspective of students, assessment is often not seen as part of the learning process, much less as a learning tool. From a functional standpoint in their experience, it makes sense why students approach new content with the mentality of, “Do I have to remember this after the test?” As such, concentrated testing experiences actually discourage learning anything that is unlikely to be tested. Students view assessment as the end of the journey for the specific material being covered.

Sound psychology programs should demand more than surface learning. Psychology as a discipline requires critical thinking, problem solving, and the application of concepts; scientific inquiry is the foundation of psychology. Attention to building those skills is essential, but that focus can get lost when students are busy with flash cards and other shallow learning practices.

**The Battery of Multiple-Choice Testing**

For students, a majority of undergraduate assessment experiences involve traditional objective testing methods such as multiple-choice questions (Fulcher, 2014). Because multiple-choice tests measure some types of learning efficiently, their use is widespread throughout the world. Although it is arguably timesaving for the teacher, multiple-choice testing has some adverse effects. For example, Saljo (1975) found that questions based on factual information from content, such as multiple-choice questions, encouraged surface learning. Students who prepare for these types of assessment often demonstrate less conceptual knowledge of the material than students who prepared for reasoning questions (as cited in Struyven et al., 2005). Success on multiple-choice testing simply does not require deep learning. Perhaps something about the format of the multiple-choice test makes it easy to dismiss once the experience is over. In addition, multiple-choice exams could potentially create false knowledge if students...
select wrong answers and do not have the opportunity to determine whether their conclusions were on target or not (Roediger et al., 2005). Rarely will students review what they did not understand when presented with a graded exam unless a teacher requires that additional step.

Teachers may inadvertently collude with students to go shallow by developing study guides that mirror the content of the multiple-choice test. Some study guides mimic the actual test so closely that the study guide feels like reworded test questions. Such practices endorse a piecemeal approach to learning.

Multiple-choice questions simply cannot always capture the breadth or depth of learning. The APA Guidelines for the Undergraduate Psychology Major: Version 2.0 (APA, 2013) promote skill development as critical to student success, yet it is hard, if not impossible, to demonstrate a variety of improved skills through multiple-choice testing.

**The Erosion of Enthusiasm From Excess Testing**

Students of this era demonstrate the impact of being part of the “No Child Left Behind” generation, a mandate established in 2001 by President George W. Bush. As such, they have experienced getting “assessed” at every opportunity. It is no wonder that students might have a hard time mustering enthusiasm for college-level assessment purely on the basis of being burned out from so much scrutiny. Constant testing simply undermines the intrinsic rewards of learning.

To complicate matters, scheduling assessment activities during a regular college term tends to make multiple assessment obligations overlap in a few high-intensity intervals (i.e., midterms, finals). All instructors in one educational setting share the same time frame; it is unavoidable that students find they must be simultaneously preparing to do their best on four to five assessments. However, such intensity is unlikely to showcase student learning at its best.

**The Perils of the Least Powerful Stakeholder**

The student is far removed from the discussions and decisions that take place between government agencies and accrediting bodies, and more often than not, students are not a part of the discussions about assessment at their institutions, within their program’s department, or even in their own classes. The lack of student participation in assessment design means that teachers typically do not benefit from students’ input on what kinds of assessment might be most helpful in demonstrating their learning. At the very least, instructors should explain the purpose of assessment and the role the data may play in making decisions that will affect the major.

In the case of assessing program quality, students may not be aware of the uses that derive from their individual assessments to measure how their programs and professors are performing. If assessment results are going to be used to make big decisions, and these results are determined by the student, it seems reasonable to gather input from the students as partners in the process. Providing well-timed feedback on performance and asking for feedback on how assessment strategies might be improved could be a dynamic process between instructors and students, allowing students to improve their learning processes and instructors to improve their teaching.

**Empowering Students as Stakeholders**

A broad plan of action for how the education and assessment experience can be improved to empower students in the process is to concentrate on consistency, continuity, and collaboration. Consistency is achieved through unambiguously defined expectations. In the APA Guidelines for the Undergraduate Psychology Major: Version 2.0, the APA (2013) established clear standards for the undergraduate psychology program. When these guidelines are used to design program curricular maps, we can facilitate greater consistency and continuity throughout the major. Sharing goals with students—and even engaging students in a discussion about how they can best meet those goals—develops a broader picture of what they should be learning and what skills they should be acquiring throughout their undergraduate career. Student voices—especially the voices of students from minoritized and stigmatized populations—must
be part of the conversation for assessment to be equitable and inclusive (McArthur, 2017). Engaging students in a conversation about the goals of assignments, eliciting how they relate to a particular learning objective, and even inviting their input on the criteria to be assessed will develop a mutual understanding of educational goals and eliminate the need for students to ask their instructors, “What do you want on this project?” However, this forces educators to confront their own beliefs and biases about what constitutes evidence of learning.

Rather than only taking up space in the syllabus, student learning outcomes (SLOs) should be few in number and meaningfully integrated throughout a course. At the beginning of the semester, SLOs give students an idea of which direction they should be going; at the end of the semester, they allow students to review what has been covered and what skills they should have gained, to recall and consolidate a term’s worth of information, and to consider their personal development and growth as a consequence of engaging with ideas in the course. Along the way, low-stakes, formative assessments engage students and mark their progress. Authentic assessment strategies connect what they are learning to their lived experience in meaningful ways (Mueller, 2018).

Beyond being uninformative and unempowering to students, the kinds of perfunctory assessment practices described throughout this review have a uniquely adverse impact on students of psychology. If we, as instructors, avoid or engage in sloppy measurement strategies in our accountability practices, we undercut the lessons we teach regarding carefully creating operational definitions of constructs and critically examining and interpreting data. Enacting thoughtful assessment strategies provides an opportunity to model the very heart of what we do best—making sense of behavior in an empirical context.

As if that is not reason enough for psychologists to critically examine their own assessment practices, there is an even more compelling impetus. As we enter an era in which social justice takes center stage, systemic disparities among our students are undeniable and intolerable. As educators, we must examine our perspective on assessment strategies and whether our practices contribute to closing or maintaining gaps among our student population (Mon-tenegro & Jankowski, 2020). All assessment is situated within a cultural context; as educators, we must ask ourselves what culture we are being responsive to in our assessment practices. Is assessment merely a response to the demands of legislators, accreditors, or administrators? If we are centered on our students and their learning, assessment tools and practices must be relevant and responsive to the needs of students—this will naturally vary by institution and even within an institution by student population (McArthur, 2017). In the context of traditional power hierarchies, assessment can systematically privilege some students and marginalize others. This works at cross purposes with the goals of higher education to create opportunity and promote equity. Assessment tools and evidence must evolve as our student populations have; to be culturally responsive to diverse students, our assessment tools and evidence must be equally diverse. As educators, we must ask whether our assessment practices help us to understand why students are achieving and whether we are empowering students to understand this for themselves. It is our hope that such social justice considerations will influence and permeate conversations about assessment for years to come.

Pedagogical Solitude Versus Communities of Practice

In conclusion, we offer teachers four general principles of the opportunities that thoughtful assessment programs can provide to improve student learning and build an inclusive, responsive assessment culture:

1. **Intentional design.** Too frequently, assessment strategies are determined in the absence of thoughtful planning, reflection, and careful design. Course design should precede the design of assessments and should be as intentional as the research-design process. Assessment should fulfill a meaningful role for the people involved.

2. **Culturally responsive strategies.** True appreciation for the diversity of students demands that assessment strategies be equally varied and diverse. Assessment is not an objective process. It is influenced by individuals’ beliefs and expectations as well as the norms and culture within an institution. Culturally responsive assessment is context dependent; it requires
“practices which respond to the needs of the contexts in which we teach and learn; including the needs of the students we serve” (Montenegro & Jankowski, 2020, p. 6). By engaging students in conversations about assessment and drawing upon multiple forms of evidence, we avoid privileging certain ways of demonstrating knowledge and promote equity and inclusion.

3. Collaboration. Almost 30 years ago Shulman (1993, p. 6) decried the “pedagogical solitude” that has long plagued teaching and called for teachers to make their work more visible within their communities. When members of an academic community view themselves as an instructional team, assessment practices can be integrated and sustained to meaningfully support student learning. When teaching is viewed as a communal effort, instructors can align their courses to one overarching curricular framework to benefit all stakeholders.

4. Transparency. If we want to be transparent in our assessment of student learning, we also must be transparent in our teaching. Teaching is too often the focus of critique and too infrequently the focus of development. Teaching benefits from collaborative, formative peer feedback. An encouraging trend in higher education is seen in “open teaching weeks” where instructors can observe colleagues for the benefit of their own teaching development, to learn new skills and approaches (e.g., Mueller & Schroeder, 2018). In such a framework, observation is an opportunity to learn from one another and develop new skills, not a performance for evaluation.

When programs embrace these principles in teaching, they can function as a community of practice, in which individuals learn from others in their community and gradually adopt shared practices (Hoadley, 2012). Instructors can share their teaching experiences with one another formally and informally, trying new teaching practices together and sharing the results within the community and sharing decision making about courses, curricula, and assessment (Beers et al., 2020). To be truly inclusive, communities should also embrace and empower students to take ownership of their learning and to contribute to discussions about learning outcomes and assessment.

Assessment is at a crossroads. The public, accreditors, administrators, teachers, and students all have a stake in the process and stand to benefit from well-designed and well-executed assessment practices. Yet to date, the promise of assessment in psychology has not been fulfilled. We have more to do to understand what our students are learning, to energize improvements to the teaching and learning process, and to support equity and inclusion. If we can remember what we are called to do and maintain a focus on learning as our prime directive, only then will assessment fulfill its true purpose in psychology.

References


Fulcher, G. (2014). *The multiple-choice test: Truly objective assessment?* [Website]. https://www2.le.ac.uk/professors/gfulcher/multiple-choice


Received July 11, 2020
Revision received August 31, 2020
Accepted September 29, 2020

---

**E-Mail Notification of Your Latest Issue Online!**

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at [https://my.apa.org/portal/alerts/](https://my.apa.org/portal/alerts/) and you will be notified by e-mail when issues of interest to you become available!