Student-Level
ASSESSMENT

ETS Research Report Series

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

Continuing a Culture of Evidence: Student-Level Assessment

Ross Markle

Educational Testing Service, Princeton, NJ

From 2006 to 2008, Educational Testing Service (ETS) produced a series of reports titled *A Culture of Evidence*, designed to capture a changing climate in higher education assessment. A decade later, colleges and universities already face a new set of challenges resulting from societal, technological, and various other influences. ETS is now initiating a new series of reports that describe current issues and trends in assessment and how they relate to the needs of higher education institutions. In this report, I discuss an increasing need to understand individual students rather than whole institutions. Previous models of assessment, which focused on comparability and accountability, made inferences at the group level, often comparing one institution to a group of similar institutions. However, emerging trends in the ways students learn, the sources of the learning, and the use of assessment results have made aggregate data alone insufficient. Greater consideration of student-level data could help to foster emerging educational trends as well as guide learning in existing colleges and universities.

The original *A Culture of Evidence* series (Dwyer, Millett, & Payne, 2006; Millett, Payne, Dwyer, Stickler, & Alexiou, 2008; Millett, Stickler, Payne, & Dwyer, 2007) was designed to capture critical trends and inform quality practice in higher education assessment. In the relatively brief time since those reports, however, new trends have emerged, and new factors have influenced both the higher education climate and the role of assessment therein.

Those reports described great variance within both institutions and students (which still holds true), and the need for better assessment at that time was driven by two key factors. First, because of this great variance in students and institutions, evaluating colleges and universities was incredibly difficult. For example, consider individual students attempting to decide which schools they would attend. Depending on their backgrounds, skills, and goals, a variety of colleges could meet their needs. Similarly, the various constituencies seeking to evaluate institutional effectiveness (e.g., accreditors, state agencies, policy makers) were trying to compare a broad and diverse space. Regardless of the metrics used, institutions can be expected to vary based on their characteristics, missions, and student populations. Thus quality assessment was (and still is) needed to accommodate a very heterogeneous collection of institutions.

The second reason to focus on assessment dealt with the types of both student and institutional data used in these evaluations. The authors of the previous reports noted that nearly all data used to compare institutions focused on either inputs (e.g., average admissions test scores) or outputs (e.g., graduation or employment rates), with little consideration for the processes of instruction and learning that take place in between. The third report (Millet et al., 2008) in the earlier *A Culture of Evidence* series went on to present a model of assessing student learning that would address these two issues and meet the need to compare institutions in a more relevant manner.

Although the need for inferences at the institutional level still exists, there are also fundamental differences that affect the way assessments are developed, delivered, and used due to the increasing need to also gather information at the individual student level. Interestingly, the need to assess individual students has not come about sharply or from a single cause. Several phenomena have emerged to create the need for more student-level information. Here I describe these phenomena in three broad categories: the source of education, the model of education, and the uses of assessments.

Corresponding author: R. Markle, E-mail: rmarkle@ets.org

Changes in the Higher Education Landscape

A Change in Educational Sources

In the 1990s, it became apparent, based on a steadily growing body of research, that attendance patterns in higher education had changed. No longer did students simply enroll in a college or university, stay for 4 years, and graduate with a degree. Students very often transferred out of and into institutions, alternated their enrollment status between full and part time, and changed majors or degree goals. In his hallmark report, *Answers in the Toolbox*, Adelman (1999) stated that "the increasing complexity of attendance patterns is one of the most significant developments in higher education of our time" (p. 38). In short, varying attendance patterns represented a revolution in the way that students experienced and used higher education.

As it has with many other things, the internet has brought yet another revolution. In addition to attending multiple institutions, students now obtain knowledge, skills, and training from extra-institutional resources such as open-educational resources (OERs) and massive open online courses (MOOCs). MOOCs generally mimic a traditional course structure, including a definitive schedule, faculty leadership, and assignments. Students frequently receive an acknowledgment of course completion and, in some cases, credit that can be applied at the sponsoring institution. OERs, however, represent a more fluid and less traditional educational resource. Often in the form of video lectures (e.g., the Kahn Academy, iTunes University), they are not tied to a course or formal structure but can generally be accessed for the purposes of supplemental instruction or sheer curiosity.

There are several reasons why institution-level assessment has less bearing in these settings. For one, these resources hold little or no tie to the institutions that are so often being compared and evaluated. Even when offered by traditional colleges and universities, many resources are not used by traditional, degree-seeking students. A survey by *New Republic* found that only 13% of MOOC participants were seeking knowledge toward a degree, and 86% of American MOOC students already had a college degree (Alcorn, Christensen, & Emmanuel, 2014). What is more, these resources still exist as a relatively new medium of education and thus are less susceptible to the calls for accountability directed at brick-and-mortar colleges and universities (Straumsheim, 2013).

Perhaps most importantly, these models of education continue the trend that started with the varying enrollment patterns realized in the 1990s. Higher education is no longer a linear pathway. Students seek higher education for many reasons, at many times, and in many capacities. A large portion of the student population still consists of 18- to 22-year-old recent high school graduates who are at the beginning of their careers, but there are also existing workers returning for retraining or perhaps coming to college for the first time—both seeking the skills to improve their career potential or, as the OER movement clearly represents, individuals who are intrinsically interested in learning.

At this point, it is important to clarify that much of the discussion on the role of MOOCs and OERs is conjecture about who is using these resources and how they are being used. Liyanagunawardena, Adams, and Williams (2013) conducted a review of research on MOOCs, finding that the number of articles, reports, conferences, and workshops on MOOCs grew from one in 2009 to 24 in 2012, yet they ultimately concluded that "the peer-reviewed research literature on them is growing but still limited" (p. 227). Athought they reported that many studies provided demographic information about students, they provided no synthesis of these findings. What is more, Williams (2013) called for a systematic census of MOOC students to inform planning, teaching, and research.

With such variant paths into and through higher education, as well as variant sources of learning and a lack of understanding about who is using them and how, there is a greater need to assess individual students rather than institutions. One reason is that single institutions are no longer the sole educator for most students. Whether this is due to more traditional factors, such as transfer behavior, or these more novel sources of education, many students no longer acquire knowledge and skills from just one place. Consider a student who enrolls at a traditional 4-year college and graduates 4 years later from that same institution. If that student were assessed upon entry and once again upon graduation in a skill such as information literacy, one could certainly make inferences about the gains in knowledge or skills that this individual student made during that time. However, to infer that these gains could be entirely attributed to that 4-year college could be erroneous. What if a student transferred to a community college for 1 year and then returned to the original school? What if the student's education was supplemented by the use of online lectures or a free MOOC? In these cases, the student certainly gained information literacy skills over that 4-year period, but to attribute those gains solely to the institution would not be valid.

Accordingly, student-level assessments also become increasingly necessary to allow individuals to better demonstrate their achievements. With multiple avenues for learning, students also experience a greater need to demonstrate their skills. For so-called nontraditional students, this has been a perennial issue, and the field of prior learning assessment (PLA) has attempted to translate the years of training and experience that some students acquire through their careers into college credit. Traditional students using OERs and MOOCs to acquire knowledge and skills may, at some point, seek to transfer evidence of that learning into another form of educational currency, such as college credits or employability credentials (e.g., Straumsheim, 2015).

Changes in Higher Education Models

The need for greater student-level assessment has also arisen due to fundamental changes in some educational models. Competency-based education (CBE) continues to gain favor because of its multiple perceived advantages. One benefit of CBE models is that they are not time bound (Johnstone, Ewell, & Paulson, 2010), allowing students to complete a section of the curriculum upon demonstrating a particular competency—even if they can already do so upon enrollment. Accordingly, CBE models are viewed as better suited for nontraditional or adult learners who have significant applied experience. Second, CBE models are more student driven (Johnstone et al., 2010), as students are generally provided with several potential tasks and charged with determining how best to complete them. Although instructional and human resources are available for additional support, students are able to determine the number of tasks needed to acquire the competency and the speed at which they can be completed. These features align CBE models with students' learning styles and profiles while enabling them to move forward at their own pace.

Taken together, the student-driven focus of CBE models and their potentially shorter duration lead to their third advantage: lower cost. Although higher education has struggled to control the price that is ultimately paid by students, CBE models have a fundamentally different structure so that the total cost—and not just the price paid by the student—is decreased. Many feel that this decreased cost underlies the Department of Education's push (Duncan, 2011; White House Office of the Press Secretary, 2013) for the adoption of CBE programs (cf. Neem, 2013).

Despite these advantages, a central challenge facing CBE models is associated with their most integral component: the demonstration of competencies. Put differently, questions arise as to which competencies shall be assessed, and how. As the Council for Adult and Experiential Learning (CAEL, 2013) stated, "Learning … can come from just about anywhere. It is the assessment of whether or not the learning the student demonstrates reflects the learning necessary for the academic credential being sought that matters most" (p. 9). The council's report also points out that an array of approaches may be used to demonstrate learning, including using standardized measures or performance tasks, real or simulated work products, or a portfolio of tasks or materials. Regardless of whether a student enrolls in a strictly CBE program, or if CBE components are adopted into traditional seat-time models, the growth of CBE will drive the need for quality, student-level assessments.

Changing Uses of Assessment

The proliferation of multiple sources of learning and the adoption of new CBE models reiterate the need to focus on assessment as a central vehicle to certify student-level achievement or completion. Of course, the use of assessments to certify course completion is not new and has occurred for decades in programs such as PLA and, since the late 1960s, in the College Board–administered College Level Examination Program (the *CLEP*® examinations), which provides students the option to earn credit for a range of courses. The final exams given by most faculty to earn course credit also serve as an often overlooked example of assessment for achievement. Generally, a minimum grade is required on these assessments for course completion and to receive course credit in partial completion of the requirements needed to earn a degree. Whereas course exams functioned as the modus operandi within an institution, CLEP allowed the communication of this completion across institutions.

The increase in innovative sources of learning and the number of students utilizing them has resulted in an elevated need for student-level assessments. The reasons are related not only to acknowledging the earning of credit but also to the potential portability of one medium or context of learning (e.g., MOOCs) to another, such as a traditional university (CAEL, 2013; Klein-Collins, 2012; Johnstone et al., 2010). Moreover, progress made in CBE models will ultimately need

to be translated into credits—and vice versa—which again will require quality assessments that can communicate what a student has achieved.

Assessments can also help communicate achievement outside of educational institutions and potentially address the needs of employers and the workforce to receive tangible evidence certifying students' knowledge, skills, and other competencies. In a companion report in this series, *Expanding Skills in Higher Education* by Maria Elena Oliveri and Ross Markle, the authors discuss the increased expectations of employers with regard to the skills of college graduates, and just as assessment can be used to communicate achievement between institutions, it could also be used to communicate achievement to employers. As one example, the ACT WorkKeys assessment has been used to indicate workforce readiness for several decades (see Hendrick & Raspiller, 2011; Swaney, Allen, Casillas, Hanson, & Robbins, 2012).

Lastly, the interest in the use of assessments is not only for summative purposes but also for formative uses. Here students are given better feedback on their performance as they progress through their learning, rather than simply proclaiming success or failure at the end of a course or degree program. Yorke (2003) noted that the use of formative assessment in postsecondary settings was promising but still required a better understanding of both pedagogical and learning processes before it could be implemented. However, one study identified several positive applications of formative assessment, with the authors suggesting that formative assessment encourages students to become proactive, rather than reactive, agents in the learning process (Nicol & Macfarlane-Dick, 2006).

Each of these uses shows that, in contrast to the model presented in the previous *A Culture of Evidence* reports, higher education can no longer solely rely on institution-level assessment or aggregated findings. Increasingly, individual students can benefit from or even require information on their own performance as they become more active agents in their own learning. Interestingly, these uses also suggest that it is not simply which measure is used or the level at which the data are collected but also the way in which information is reported that must be considered.

Implications

Implications for Colleges and Universities

Student-level assessment has the potential to greatly benefit colleges and universities as they continue the use of assessment results for improvement rather than solely for accountability (Ewell, 2009; Kuh et al., 2015). Assessments used to make institutional rather than individual inferences—which are inherently low stakes for test takers—have always raised concerns about student motivation and, subsequently, the validity of score-based inferences (Thelk, Sundre, Horst, & Finney, 2009). Research by Liu, Bridgeman, and Adler (2012) demonstrated that varying levels of motivation can affect student scores and suggested that scores are related to the assessment's perceived stakes. Providing scores to individual students—as a part of the learning process, for credit completion, or for credentialing—increases the stakes of the assessment and will thereby likely increase the quality of data that institutions receive about student learning.

What is more, student-level results may have a more direct and immediate impact on students. Even in optimal improvement-based paradigms, changes to a course, program, or institutional curriculum usually take place more gradually across semesters or years. Hence students who initially participate in the assessment are typically unable to see the improvements, and only future cohorts benefit by receiving instruction informed by the assessment results. If students receive relatively immediate results on their performance, they can identify particular areas of deficit and seek remediation more swiftly. This shift would impact individual students directly.

However, there is one drawback for institutions as they move toward using more student-level data. Miller and Leskes (2005) framed various levels of data collection and the uses thereof and noted that data at the institutional level can often be sampled from larger groups of students (rather than assessing each student individually). Obviously, increasing the amount of information reported to individual students will require greater time and attention to assessment in order to produce reliable and valid scores at that level. This will also have a likely impact on institutional assessment policies, costs, and resources.

Implications for Researchers and Assessment Developers

Whereas institutions will likely need to become more adept at *using* student-level information, researchers and assessment developers will need to better understand how they can *produce* the types of student-level information that students,

faculty, and institutions seek. Even though many of the uses of student-level information (e.g., certifying completion, credentialing) are summative, lessons about how to report scores to students in transparent and meaningful ways are perhaps best learned from the formative assessment literature.

In a review of the research into online formative assessment in higher education, Gikandi, Morrow, and Davis (2011) found that self-assessments, discussion forums, and e-portfolios are the most useful approaches. Similarly, research by Hart Research Associates (2015) also found that having access to an e-portfolio summarizing and demonstrating individual accomplishments in key skill and knowledge areas (e.g., effective communication, knowledge in the field, applied skills, evidence-based reasoning, ethical decision-making) would be the most helpful means—even beyond a college transcript—to evaluate a student's potential to succeed. They also found that the use of these tools fostered student engagement in the learning process, both individually and socially (i.e., through development of learning communities).

Although tools such as these can be valuable for local and formative assessment efforts, the summative purposes identified here will require more formal, and in many cases standardized, measurement in order to be valid, secure, and portable. A key consideration for new assessments could be a move to criterion-referenced scoring. Criterion-referenced scoring indicates learning in reference to an established benchmark of performance and can be contrasted with normative scoring, which refers to a student's position relative to other students. Criterion-referenced scoring can facilitate formative assessment and subsequently the improvement of student learning (e.g., O'Donovan, Price, & Rust, 2001; Rust, Price, & O'Donovan, 2003; Wilson & Scalise, 2006). By conveying information in relation to learning goals, students and institutions alike can better understand what has been learned and not just rankings among peers.

Implications for Public Policy

To some extent, public policy has already addressed one of the key issues related to student-level assessment. The Higher Education Reconciliation Act of 2005 contained the direct assessment provision, which allowed students enrolled in programs that acknowledge achievement in terms of performance (rather than credit hours or seat time) to receive major forms of federal financial aid. However, due to some ambiguity about the provision, it was not until a 2013 clarification by the U.S. Department of Education that the viability of competency-based programs became clear (Muir & Goldstein, 2013; Porter, 2014). Ultimately, this policy acknowledged the importance of student-level assessment in demonstrating achievement of learning and completion of a program.

Moving forward, however, policy makers will need to monitor the use of the emerging technologies and models discussed here. MOOCs and OERs hold great value at present by providing intrinsic and multifunctional learning opportunities; that is, individuals can participate in a MOOC to sharpen skills, support other forms of formal instruction, or simply because they wish to learn about the topic. However, if those roles change (e.g., increased instances of students seeking credit for MOOC participation) or MOOCs become a larger part of traditional institutions' operations, then accountability—and, accordingly, assessment—will have increased relevance in this space.

Conclusions

When developing new assessments or when seeking to understand the validity of existing measures, the central issue is the use of assessment results (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Kane, 2001). We have outlined several uses here—indicating achievement, fostering improvement, credentialing—that focus on the individual and that have all essentially emerged since the previous *A Culture of Evidence* series concluded in 2008. These rapid changes have left higher education seeking to understand how it will operate in this climate. Some have even posed that such fundamental shifts in the way we measure and certify learning could threaten the postsecondary system's very existence (Christensen, Horn, Caldera, & Soares, 2011).

Whether these changes significantly impact higher education as a whole, they will certainly affect both the uses and the users of assessment results. Subsequently, they will certainly impact assessments themselves. The types of measures used, the way they are implemented, the types of data reported, and the mechanisms for reporting scores are all likely to change to meet the need to better understand learning at the individual student level.

Ultimately, however, many constituents of higher education should view this change as a positive step. An increased desire to assess individual students represents, it is hoped, an increased interest in student achievement—a focus on

individual learning rather than just instruction to large groups. As a greater number of students access a greater number of institutions for an increasing number of reasons, it will be critical to understand their unique paths, successes, and ultimate outcomes. Of course, assessment will be a key vehicle for gaining that understanding.

References

Adelman, C. (1999). *Answers in the tool box: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education.

Alcorn, B., Christensen, G., & Emmanuel, E. J. (2014, January 4). Who takes MOOCS? *New Republic.* Retrieved from http://www .newrepublic.com/

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Christensen, C. M., Horn, M. B., Caldera, L., & Soares, L. (2011). *Disrupting college: How disruptive innovation can deliver quality and affordability to postsecondary education.* Washington, DC: Center for American Progress.

Council for Adult and Experiential Learning. (2013). Competency-based education. *CAEL Forum and News*. Retrieved from http://http://www.cael.org/pdfs/cael_competency_based_education_2013

Duncan, A. (2011, November 29). *Beyond the iron triangle: Containing the cost of college and student debt*. Remarks of U.S. Secretary of Education Arne Duncan to the annual Federal Student Aid Conference, Las Vegas, NV. Retrieved from http://www.ed.gov/news/ speeches/beyond-iron-triangle-containing-cost- college-and-student-debt

Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes*. Princeton, NJ: Educational Testing Service.

Ewell, P. T. (2009). *Assessment, accountability, and improvement: Revisiting the tension* (National Institute of Learning Outcomes Assessment Occasional Paper No. 1). Retrieved from http://www.learningoutcomesassessment.org/documents/PeterEwell_005.pdf

Gikandi, J. W., Morrow, D., & Davis, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*, 2333–2351.

Hart Research Associates. (2015). *Falling short? College learning and career success*. Washington, DC: Association of American Colleges and Universities. Retrieved from https://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf

Hendrick, R. Z., & Raspiller, E. E. (2011). Predicting employee retention through preemployment assessment. *Community College Journal of Research and Practice, 35*, 895–908.

Johnstone, S. M., Ewell, P., & Paulson, K. (2010) *Student learning as academic currency*. Washington, DC: American Council on Education.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*, 319–342.

Klein-Collins, R. (2012). *Competency-based degree programs in the US: Postsecondary credentials for measurable student learning and performance*. Chicago, IL: Council for Adult and Experiential Learning. Retrieved from http://files.eric.ed.gov/fulltext/ED547416 .pdf

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

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education motivation matters. *Educational Researcher, 41*, 352–362.

Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs: A systematic study of the published literature 2008–2012. *International Review of Research in Open and Distributed Learning, 14*, 202–227.

Miller, R., & Leskes, A. (2005). *Levels of assessment: From the student to the institution*. Washington, DC: American Association of Colleges and Universities.

Millett, C. M., Payne, D. G., Dwyer, C. A., Stickler, L. M., & Alexiou, J. J. (2008). *A culture of evidence: An evidence-centered approach to accountability for student learning outcomes*. Princeton, NJ: Educational Testing Service.

Millett, C. M., Stickler, L. M., Payne, D. G., & Dwyer, C. A. (2007). *A culture of evidence: Critical features of assessments for postsecondary student learning*. Princeton, NJ: Educational Testing Service.

Muir, G., & Goldstein, M. B. (2013). Establishing federal financial aid eligibility for competency-based education. *Career Education Review*.

Neem, J. N. (2013). Experience matters: Why competency-based education will not replace seat time. *Liberal Education, 99*(4), 26–29.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*, 199–218.

O'Donovan, B., Price, M., & Rust, C. (2001). The student experience of criterion-referenced assessment (through the introduction of a common criteria assessment grid). *Innovations in Education and Teaching International, 38*(1), 74–85.

Porter, S. R. (2014). *Competency-based education and federal student aid*. Retrieved from http://www.cbenetwork.org/sites/457/uploaded/files/Competency_Based_Education_and_Federal_Student_Aid.pdf

Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education, 28*, 147–164.

Straumsheim, C. (2013, December 19). Go easy on MOOCs. Inside Higher Ed. Retrieved from http://http://www.insidehighered.com/

Straumsheim, C. (2015, April 24). Change, but how substantive? *Inside Higher Ed.* Retrieved from http://www.insidehighered.com/

Swaney, K. B., Allen, J., Casillas, A., Hanson, M. A., & Robbins, S. B. (2012). Interests, work values, and occupations predicting work outcomes with the WorkKeys Fit assessment. *Journal of Career Assessment, 20*, 359–374.

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Motivation matters: Using the Student Opinion Scale to make valid inferences about student performance. *Journal of General Education, 58*, 129–151.

White House Office of the Press Secretary. (2013). *Fact sheet on the president's plan to make college more affordable: A better bargain for the middle class*. Retrieved from http://www.whitehouse.gov/the-press-office/2013/08/22/fact-sheet-president-s-plan-make-college-more-affordable-better-bargain-

Williams, B. (2013). *Roll call: Taking a census of MOOC students*. Paper presented at the 1st Workshop on Massive Open Online Courses at the 16th annual Conference on Artificial Intelligence in Education, Memphis, TN. Retrieved from http://hwww.moocshop.org/

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. *Higher Education, 52*, 635–663.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*, 477–501.

Suggested citation: