1

Indicators of Educational Quality for Postsecondary Accountability Systems

Aaron S. Horn and David A. Tandberg

The United States is widely revered for its presumed excellence in higher education, and it now seems commonplace to characterize our postsecondary system as "the envy of the world" (e.g., Khator 2011). In fact, according to the Academic Ranking of World Universities—which heavily weights research productivity—fifty of the top 100 universities globally reside in the United States (Shanghai Ranking Consultancy 2016). Few informed observers, however, would equate research prowess with educational effectiveness. Indeed, the educational quality of postsecondary institutions has been increasingly called into question as evidence periodically surfaces of marginal knowledge and skill acquisition among college graduates (Arum and Roska 2011; Kutner et al. 2007; Desjardins et al. 2013). For example, results from the National Assessment of Adult Literacy indicated that only 40 percent of bachelor's degree recipients were proficient in prose literacy in 1992, a figure that fell to 31 percent in 2003 (Kutner et al. 2007). More recently, a national survey revealed that only 24 percent of employers agreed that college graduates were well prepared for "applying knowledge/skills to [the] real world" (Hart Research Associates 2015, 12).

Concerns about educational quality have also emerged amidst competing state budgetary priorities and low graduation rates, as colleges and universities are pressured to provide evidence of their value, effectiveness, and efficiency, thereby ensuring that taxpayer investments

Assessing Quality in Postsecondary Education: International Perspectives, edited by Harvey P. Weingarten, Martin Hicks, and Amy Kaufman. Montréal and Kingston: McGill-Queen's University Press, Queen's Policy Studies Series.

in higher education are being utilized responsibly. Policy makers have thus issued a clarion call for greater accountability related to educational quality in the postsecondary sector, particularly the need to collect and publicize data on student learning outcomes (e.g., Reindl and Reyna 2011; SHEEO 2005; US Department of Education 2006). This chapter seeks to inform policy discourse by providing an overview of performance indicators used to evaluate educational quality for purposes of public accountability and improvement in higher education.

Toward Which End?

If the old adage that *whatever is measured is valued* holds true, a comprehensive conception of the aims of higher education must remain at the fore to ensure that stakeholders understand not only what is valued in student learning, but more importantly, what is not. While the public is generally aware that colleges should promote degree completion to the extent possible, few students enter (or leave) college with an understanding of what should be learned. Nonetheless, much ink has been spilled over attempts to delineate essential learning outcomes, frequently with remarkable commonalities across frameworks (e.g., AAC&U 2007; Adelman et al. 2014; Markle et al. 2013). Three national initiatives to establish norms for student learning outcomes are noteworthy. In its report, *College Learning for the New Global Century*, the Association of American Colleges and Universities (2007) outlined a set of student outcomes that were endorsed by a panel of diverse leaders representing the education, business, non-profit, and government sectors. The panel sought to identify the critical competencies for life, work, and citizenship that should be expected of all college graduates, regardless of degree level or institution type. The resulting learning outcomes were subsumed within four rubrics: (a) knowledge of human cultures and the physical and natural world; (b) intellectual and practical skills; (c) personal and social responsibility; and (d) integrative and applied learning (see Table 1.1).

In another prominent example, the Lumina Foundation's Degree Qualifications Profile (DQP) provided a similar conceptual framework for learning outcomes: specialized knowledge, broad and integrative knowledge, intellectual skills, applied and collaborative learning, and civic and global learning (Adelman et al. 2014). Moreover, the DQP extended previous work by specifying levels of proficiency for each outcome by degree level (i.e., associate, bachelor's, master's). Finally,

Table 1.1

AAC&U's Ideal Learning Outcomes

Outcome Rubric	Defining Elements
Knowledge of Human Cultures and the Physical and Natural World	Through study in the sciences and mathematics, social sciences, humanities, histories, languages, and the arts
Intellectual and Practical Skills	Inquiry and analysis Critical and creative thinking Written and oral communication Quantitative literacy Information literacy Teamwork and problem solving
Personal and Social Responsibility	Civic knowledge and engagement—local and global Intercultural knowledge and competence Ethical reasoning and action Foundations and skills for lifelong learning
Integrative and Applied Learning	Synthesis and advanced accomplishment across general and specialized studies

Source: Adapted from *College Learning for the New Global Century* by the Association of American Colleges and Universities (2007).

whereas the aforementioned initiatives mainly focused on learning outcomes applicable to all college graduates, the Social Science Research Council led a two-year project, Measuring College Learning (MCL), to facilitate faculty discourse on disciplinary learning outcomes (Arum, Roska, and Cook 2016). The MCL project engaged faculty in biology, history, economics, communication, sociology, and business.

Defining Educational Quality

Numerous competing definitions of quality have been used, implicitly or explicitly, in the field of higher education with varying emphases on inputs, processes, and outcomes (Harvey and Green 1993; Campbell 2015). The resources and reputation model is the oldest approach, for example, and equates quality with prestige and inputs. This model is readily observed in nearly every ranking system produced for popular consumption, such as *US News and World Report's Best Colleges*, which

evaluates institutions based on class size, faculty salary, admissions test scores, per-student expenditures, and other factors. A major problem with this approach is that the presence of impressive resources and a strong reputation among peers do not permit inferences about effective resource utilization and learning outcomes, which are central concerns in the context of public accountability. Research has long revealed only modest relationships between traditional input measures and learning gains in college (Mayhew et al. 2016), thereby preserving the etymological roots of prestige in the Latin *praestīgiae*, or juggler's tricks.

The working definition of quality in this chapter, therefore, focuses less on inputs and more on educational practices and student outcomes, namely the extent to which an institution meets reasonable standards in (a) employing programs, practices, and policies that are generally known to be conducive to student learning and timely degree completion; (b) enabling and adding value to student outcomes; and (c) ensuring that graduates have fulfilled learning objectives. These dimensions of educational practice, institutional effectiveness, and degree integrity, along with exemplary performance indicators, are elaborated below.

Good Educational Practice

Educational practice indicators have rarely been considered within state accountability systems based on the principle that institutions should retain autonomy over the design of the curriculum, pedagogy, and support services. However, this objection withers with the realization that the amount of autonomy ceded will be contingent on the level of practice specification and the type of accountability system. For instance, the specification of a good pedagogical practice such as "giving useful feedback" is sufficiently abstract to permit a myriad of qualifying faculty behaviours; there is no standard response *X* prescribed in classroom situation *A*. Moreover, a process orientation to educational quality is complementary insofar as it offers some accountability-related advantages over outcome-based assessment approaches. Whereas an institution can assume responsibility for educational practices, student-learning outcomes are partly a function of factors that frequently lie beyond institutional control, particularly student academic aptitude, preparation, and motivation (e.g., Liu, Bridgeman, and Adler 2012). Even when such extraneous influences are parsed out of the outcome in the form of an adjusted gain score, institutional leaders and faculty are left with little actionable knowledge for improvement. Conversely, an account-

ability system that incorporates process measures creates stronger expectations and clearer feedback for shaping faculty and staff behaviour. Finally, a central rationale for directly evaluating educational practice is the difficulty in adequately assessing the full range of ideal learning outcomes due to constraints in resources and the availability of appropriate measures. It is much easier to determine whether a putative best practice is prevalent than to ascertain its intended effect.

The identification of good educational practices should be based on expert consensus and empirical linkages with desirable student outcomes. Accordingly, this section draws on past conceptions of good practices that have been associated with academic engagement, learning, and persistence outcomes (e.g. Ewell and Jones 1996; Jankowski 2017; Kuh et al. 2011; Kuh, 2008; National Research Council 2013; Tinto 2012). For instance, among the most influential frameworks, Kuh et al. (2011) argued that highly effective institutions can be distinguished by their level of academic challenge, use of active and collaborative learning, student-faculty interaction, student involvement in enriching educational experiences, and a supportive campus environment. Kuh's framework has gained national prominence through the administration of the National Survey of Student Engagement, or NSSE (McCormick, Kinzie, and Gonyea 2013). Although far from perfect, the NSSE currently provides the best balance among key test selection criteria, specifically the cost of administration, the ease of analysis and interpretation, widespread adoption, and reliability and validity. In fact, the NSSE scales have been recently used as educational practice indicators for public accountability in Maine, Tennessee, Vermont, Wisconsin, and Wyoming. However, rather than simply convert the entire set of NSSE items into performance measures, we recommend a subset of four practice indicators that have garnered the strongest support in empirical studies: (1) instructional excellence, (2) highly effective programs, (3) academic challenge, and (4) academic and social support.

Indicator 1: Instructional Excellence

The first indicator assesses the extent to which faculty are using effective instructional techniques. For example, the National Research Council (2013) reviewed research in educational psychology and identified several pedagogical practices that are conducive to deep learning,[1] such

1 Deep (or deeper) learning refers to meaningful learning that promotes both retention and transfer of knowledge to novel situations (Mayer, 2011).

as representing concepts in multiple modes; using examples, cases, and models; and providing formative assessments that improve the learning process. More generally, after controlling for student background characteristics, researchers found that the presentation of a course in an organized and well-planned manner was positively associated with gains on standardized tests of reading comprehension (Bray, Pascarella, and Pierson 2004) and critical thinking (Loes, Salisbury, and Pascarella 2015) as well as the likelihood of student persistence (Pascarella, Salisbury, and Blaich 2011). This research provides a satisfactory basis for measuring instructional excellence with the NSSE Effective Teaching Practices (ETP) scale, wherein students report whether their professors used examples to explain difficult concepts; taught courses in an organized way; and gave timely and useful feedback, among other techniques. The ETP scale has been positively associated with self-reported learning gains during the first year of college (Zilvinskis, Masseria, and Pike 2017).

Indicator 2: Highly Effective Programs

Highly effective programs refer to structured curricular and co-curricular activities that are grounded in sound pedagogical principles and have been demonstrated to reliably yield intended learning outcomes. Kuh (2008) identified an extensive array of "high-impact practices" that have been consistently associated with self-reported learning gains including participation in a learning community, service learning, research with faculty, internships, study abroad, and senior-year capstone courses or projects. Kuh argued that such experiences tend to demand a high degree of student effort, promote faculty and peer interactions, expose students to diverse ideas and people, incite feedback from others, and require the application and integration of knowledge. However, empirical support varies considerably for these practices, and resource constraints may prevent broad diffusion in some cases, particularly providing opportunities for students to conduct research with faculty. A further limitation of the NSSE high-impact practice index is the omission of other effective pedagogical models.

Whereas evidence is generally supportive for learning communities (Weiss et al. 2015), living-learning programs (Inkelas and Soldner 2011), study abroad (Horn and Fry 2013), and internships (Reddy and Moores 2012), less is known about the impact of culminating senior experiences. In contrast, a rather robust corpus of research has demon-

strated the effectiveness of community service and service learning in both secondary and postsecondary sectors (Horn 2012; Yorio and Ye 2012). Yorio and Ye (2012) conducted a meta-analysis of forty studies that examined service learning during college in relation to cognitive development (e.g., GPA, course performance, problem solving) and found an average effect size (ES) of .53, or a difference of 20 percentile points. Other meta-analyses have estimated the effect size to be closer to .31 (Celio, Durlak, and Dymnicki 2011; Warren, 2012).

Absent from Kuh's (2008) initial list are other pedagogical models with strong empirical support, including problem-based learning and co-operative learning. Problem-based learning (PBL) is a student-centred pedagogy, which uses representative problems that organize and stimulate the learning experience (Barrows 1996). In their meta-analysis of forty-three studies, Dochy et al. (2003) found that PBL was associated with significant gains in the ability to apply knowledge (ES = .66; 25 percentile points) relative to the traditional lecture method. Collaborative or co-operative learning involves small groups of students (typically two to four per group) who collaborate in a specific way to attain a common learning goal.[2] Meta-analyses have provided overwhelming support in favour of co-operative learning relative to traditional individual or competitive learning (Johnson, Johnson, and Smith 1998; Kyndt et al. 2013). Academic achievement gains approximate an effect size of .49 to .53, or 19 to 20 percentile points. The diffusion of such pedagogical models remains one of the most urgent priorities in higher education as 51 percent of faculty at four-year institutions principally employ "extensive lecturing" in most of their courses (Eagan et al. 2014).

Indicator 3: Academic Challenge

The academic challenge indicator evaluates the rigour of the college curriculum, especially whether learning activities emphasize conceptual depth and integration rather than the simple reproduction of course material (Biggs and Tang 2011; Campbell and Cabrera 2014). To be sure, in contradistinction to the metaphor of students as empty vessels that

2 In co-operative learning, the instructor ensures that (a) individual performance is intertwined with group performance; (b) each student is held accountable for his or her performance; (c) students enhance each other's learning outcomes through explanations, modelling, reinforcement, and coaching; (d) students develop requisite social skills for effective teamwork; and (e) students evaluate and improve group work processes (Johnson, Johnson, and Smith 1998).

must be relentlessly filled with facts, educational psychologists have demonstrated that students learn by actively constructing knowledge through the explanation, application, and integration of new concepts (Barkley 2009). Academic challenges of this sort can be assessed through the NSSE Reflective and Integrative Learning scale, wherein students are asked about whether they were able to connect topics across courses; link classroom learning with community problems; and connect new course material to prior knowledge, *inter alia*. Student scores on this scale have been associated with gains in critical-thinking skills, the need for cognition, and a positive attitude toward literacy (Nelson Laird et al. 2014).

Indicator 4: Academic and Social Support

The final practice indicator evaluates the effectiveness of the institution's academic and social support. Past research has revealed that a variety of academic support programs can promote student persistence, such as first-year seminars, student success courses, tutoring, and summer bridge programs (Asmussen and Horn 2014). Effective support also involves providing opportunities for social integration (Tinto 2012) and promoting psychological welfare (Francis and Horn 2017). Pertinent items in NSSE form the supportive environment scale, which assesses student perceptions of institutional support for academic success and meaningful campus activities, for instance. While controlling for potentially confounding student and institutional characteristics, an earlier version of the supportive environment scale predicted higher retention and graduation rates (Gordon, Ludlum, and Hoey 2008; Pike 2013).

Institutional Effectiveness

The current knowledge base of good educational practice can inform the development of relevant indicators, but a quality-assurance system must permit the evolution of practice and acknowledge the limits of our ability to precisely specify the nature of good practice. The second facet of our definition of educational quality thus regards outcome-based evidence for whether the institution is enabling and adding value to student learning and degree completion. Measures of institutional effectiveness are not intended to provide focused diagnostic feedback, but rather identify a general need for closer scrutiny of institutional practices (for remediation or emulation). Two indicators are proposed

to assess added value: (a) basic skills development and (b) promoting timely completion.

Indicator 5: Basic Skills Development

Any of the learning outcomes previously outlined in Table 1.1 might be subject to assessment, but several constraints will quickly temper overly enthusiastic proposals: resources are limited, the state of test development varies, some learning outcomes enjoy greater consensus, and certain skill deficits are better documented than others. The simplest and most popular approach draws upon the self-reported (or perceived) learning gains of college seniors. For example, according to national NSSE results, the majority of students believed that their institution contributed quite a bit (38 percent) or very much (47 percent) to their critical-thinking ability (NSSE 2017). However, while a student experience survey is a low-cost solution to learning assessment, it may not be a valid one. Past research has indicated that self-reported learning gains are at best only weakly correlated with gains on standardized achievement measures (Bowman 2010, 2011; Porter 2013; Sitzmann et al. 2010). Self-reported learning gains instead appear to assess student satisfaction with the academic experience (Gonyea and Miller 2011; Sitzmann et al. 2010).

The preferred alternative to self-reported learning gains is to directly measure student learning. The most advanced testing initiatives have focused on critical thinking, writing, reading, and quantitative literacy. This partly reflects a strong consensus around these abilities among faculty at four-year institutions, wherein 99 percent believe that critical thinking is an essential or very important learning goal for all undergraduates, followed by the ability to write effectively (93 percent; Eagan et al. 2014). The ETS Proficiency Profile, the ACT CAAP, and the Council for Aid to Education's Collegiate Learning Assessment Plus (CLA+) are among the most established assessments of these basic skills. In fact, Klein, Liu, and Sconing's (2009) validation study demonstrated that institution-level scores from these three measures are highly reliable and strongly inter-correlated, though there are some important differences that should be considered before selecting a test for performance reporting.

The ETS Proficiency Profile and the ACT CAAP employ a multiple-choice format in assessing reading, writing, mathematics, and critical thinking (the CAAP also includes a science module). In support

of the ETS measure, Roohr, Liu, and Liu (2016) found that the number of completed credits during college predicted critical-thinking gains. Furthermore, the ETS measure detects average gains in critical thinking (ES = .57) that are comparable to those documented in the literature using other tests for critical thinking. For instance, Huber and Kuncel's (2016) meta-analysis revealed an average gain during college of about .59 (50th to 72nd percentile). Conversely, Klein, Liu, and Sconing (2009) found that the ACT CAAP had a very wide 95 percent confidence interval for critical-thinking gains (ES = .06 to .56).

The CLA+[3] is notable for moving beyond multiple choice to so-called constructed response formats, wherein students create their own response to a prompt rather than choose from a set of predefined options. Constructed response tests appear to better ensure that students demonstrate understanding rather than simple recall. For instance, Hyytinen et al.'s (2015) analysis of responses to multiple-choice and constructed response items suggested that students with a high multiple-choice test score, but a low CLA score, had likely engaged in superficial information processing. However, it remains unclear whether the CLA+ can reliably detect gains in critical thinking, as Klein, Liu, and Sconing (2009) found that the average adjusted CLA gain score for the performance task did not significantly differ from zero.

A common objection to using standardized achievement measures is that many students may not put forth their best effort. This is a valid concern as the results of low-stakes testing are highly contingent on student motivation (Liu, Bridgeman, and Adler 2012; Finney et al. 2016). One of the few effective solutions is to raise the stakes of the test by linking it with grades or degree conferral, though this route presents several technical and political obstacles (Wise and DeMars 2005), such as faculty resistance. In another tactic, monetary incentives have been shown to increase student motivation in some cases (Duckworth et al. 2011; cf. O'Neil et al. 2005), but the cost implications arguably threaten the long-term sustainability of test administration. More promising approaches under development modify the perceived meaning and consequences of testing. Past research has indicated that motivation can be improved by ensuring that students understand the purpose of accountability testing (Zilberberg et al. 2014) and informing students that

3 The current version of this instrument, termed the CLA+, differs from the previous CLA in important ways. However, the core performance task for assessing critical thinking and written communication remains unchanged.

test results will affect their institution's reputation and perceptions of degree quality (Liu, Rios, and Borden 2015; cf. Kornhauser et al. 2014). However, simply notifying students that test results will provide valuable feedback for personal development or will be shared with faculty has no reliable effect on student motivation (Finney et al. 2016). Finally, in the absence of effective motivational interventions, design and statistical methods have been developed to detect unmotivated students to remove them from the sample (Swerdzewski, Harmes, and Finney 2011; Rios, Liu, and Bridgeman 2014).

Indicator 6: Promoting Timely Completion

A prominent aim of recent public accountability models has been to increase the production of postsecondary credentials. Many states have established college attainment goals in response to projected shortfalls in the supply of adults with college credentials (Lumina Foundation 2016). Moreover, high dropout rates reduce the state's return on investment through lost institutional appropriations and student grant aid as well as lost revenue from state income tax (Schneider and Yin 2011). Graduation rates have thus been adopted or proposed as a core performance indicator in state accountability systems across the US. Unfortunately, raw graduation rates are mainly a function of factors over which most institutions have very limited control, such as the academic preparedness of incoming students (e.g., Adelman 2006). While the goal of promoting student persistence is praiseworthy, drawing conclusions from raw graduation rates is a highly questionable endeavour, particularly in the context of performance funding (Horn and Lee 2017).

A potential solution lies in a value-added approach that estimates the difference between actual and predicted graduation rates (Horn and Lee 2016; Horn, Horner, and Lee 2017). In this statistical method, institutional conditions are considered to be conducive to timely completion to the extent that the actual graduation rate approximates or exceeds the rate that is predicted from student characteristics and other factors that cannot be reasonably placed under institutional control. For instance, Horn and Lee's (2016) value-added model for four-year colleges and universities accounted for institutional mission, control (public, private), and size; admissions selectivity; graduate student enrolment; student demographics (i.e., academic preparedness, gender, ethnicity, age, federal grant recipients); educational expenditures; urbanicity; and the size of the state's knowledge labour market. Their

psychometric evaluation of the resulting measure demonstrated that properly specified regression models can indeed yield reliable and valid estimates of institutional effectiveness. A similar value-added completion indicator is being used for performance reporting in the City University of New York system (see McDonnell et al. 2013).

Degree Integrity

In the third dimension of educational quality, degree integrity reflects the extent to which college graduates have fulfilled learning objectives, that is, whether an institution's standards for degree conferral establish a reasonable level of concordance between actual and ideal outcomes. To be sure, perceived shortcomings in the integrity of college degrees are partly responsible for the increased scrutiny of educational quality. An increasingly popular approach among policy makers is to use employment outcomes, particularly the salary of college graduates, as proxies for skill proficiency (e.g., Sparks and Waits 2011). However, it remains unclear whether earnings constitute a defensible indicator in this regard. Studies that control for institutional selectivity, academic major, and student background characteristics, for instance, suggest that college GPA has a null or small positive effect on post-college earnings (Donhardt 2004; Zhang 2008).

Although institutions bear the ultimate responsibility for degree conferral, there are contextual contingencies that may limit an institution's performance on indicators of degree integrity, particularly a reliance on local pools of incoming students and the corresponding quality of their PK–12[4] education as well as both the economic and political consequences of depressed graduation rates resulting from unattainable academic standards. Consequently, an important caveat is that performance on degree integrity indicators is more appropriately conceived as representing the effectiveness of the PK–16 system (and beyond) rather than that of the postsecondary institution in question. While mindful of this limitation, three indicators are proposed to evaluate degree integrity: (a) basic skills proficiency, (b) major field competence, and (c) civic engagement.

4 PK–12 refers to preschool and kindergarten through 12th grade of secondary school. PK–16 refers to preschool and kindergarten through the completion of a four-year baccalaureate program.

Indicator 7: Basic Skills Proficiency

Whereas the metric for basic skills development assesses the value an institution adds to student learning, an indicator of basic skills *proficiency* provides the percentage of graduates who meet or exceed a performance threshold. Data collection for the former can thus be readily applied to the latter. In addition, current testing options afford an opportunity to benchmark performance beyond national borders to evaluate the competitiveness of a state's workforce in the global arena. The most widely used measure is the OECD's Program for the International Assessment of Adult Competencies (PIAAC), which assesses adult literacy, numeracy, and problem solving in technology-rich environments. The principal challenge is to control for the source of the respondent's education as cross-state migration can confound inferences of system effectiveness.

Indicator 8: Major Field Competence

Major field coursework typically occupies half of the undergraduate curriculum in the United States, and thus a corresponding performance indicator is in order. One option is to administer the ETS subject exams if they have been endorsed by faculty and professional associations as adequate assessments of core disciplinary knowledge. Another approach draws upon licensure exam scores of recent college graduates in relevant fields, such as education, nursing, and accounting (Miller and Ewell 2005). For instance, the Minnesota Office of Higher Education (2012) reported system-level pass rates on a teacher licensure exam for the University of Minnesota (85 percent), other public universities (97 percent), and private four-year not-for-profit institutions (95 percent). An additional aspirational standard of candidate performance could be established at the 85th percentile, similar to the ETS Recognition of Excellence Award.

Indicator 9: Civic Engagement

A measure of whether college graduates become active citizens provides a final test of educational quality. Civic engagement can be defined as any behaviour that has the intent or effect of influencing matters of public interest, especially the protection, promotion, or provision of public goods and rights (cf. Levine 2007). Zukin et al. (2006) proposed one of the more extensive enumerations of relevant behaviours, including (a) civic indicators (community problem solving, volunteerism, member-

ship in or donations to an association, fundraising); (b) political indicators (voting, persuading others, displaying campaign paraphernalia, donations, volunteering for a campaign); (c) public-voice indicators (contacting officials or media, protesting, signing petitions, boycotting, "buycotting," canvassing); and (d) cognitive engagement (following government affairs and the news, discussing politics, political knowledge). We recommend that states report civic engagement frequencies based on alumni surveys or senior-year surveys, such as NSSE's topical module on civic engagement (NSSE 2017). A central purpose of this type of indicator is to signal to both institutions and stakeholders, especially incoming students, that active citizenship is an expected and highly valued learning outcome.

Utilization of Quality Indicators

The use of quality indicators in accountability systems varies considerably across the United States. An analysis of state and system accountability reports over the past two years revealed twenty states with at least one of the aforementioned indicators of educational quality.[5] This represents a decrease from earlier decades as states dropped quality indicators ostensibly due to the high cost of test administration; loss of political support; changes in leadership; a questionable impact on student learning; and the delegation of student outcomes assessment to regional and program accreditation (Ewell 2009). Most of the states in our analysis used quality indicators for performance reporting, whereas Florida, Kansas, Missouri, Pennsylvania, and Tennessee incorporated indicators within a funding model. For example, Tennessee's "quality assurance funding" model links approximately 5 percent of institutional funding with indicators derived from multiple measures, such as standardized general and major education assessments, program accreditation status, and the NSSE (THEC 2015).

Less clear is the extent to which state efforts to evaluate educational quality have affected institutional practice and student outcomes. On the one hand, state mandates, along with new requirements for regional and program accreditation, have been recognized as a strong impetus for the historical evolution of campus-based student learning assessment (Ewell 2008). In a national survey of provosts from over 600 public postsecondary institutions, respondents rated state and system

5 Interested readers should contact the authors to obtain the related content analysis table.

mandates as a moderate motivational force behind assessment activities, though accreditation was deemed the most important factor (Kuh et al. 2014).[6] Furthermore, on average, provosts indicated that learning assessment results play a moderate role in curricular development, the revision of learning goals, and strategic planning. Respondents assigned less significance to using assessment results for resource allocation and budgeting. On the other hand, rigorous studies of performance funding provide a cautionary note for those who seek to link educational quality measures with state appropriations. Specifically, the preponderance of evidence suggests that performance funding has a negative or null effect on degree productivity, except for potentially small positive effects on the production of short-term certificates (e.g., Hillman, Hicklin Fryar, and Crespín-Trujillo 2017; Rutherford and Rabovsky 2014; Tandberg and Hillman 2014). Among the unintended consequences, institutions exposed to performance funding were more likely to limit the enrolment of racial and ethnic minority students (Umbricht, Fernandez, and Ortagus 2017) and lower-income students (Kelchen and Stedrak 2016).

Despite the expanded institutional capacity for assessment, the impact of quality measurement may be limited on many campuses due to a "compliance culture" that fails to balance the formative and summative functions of evaluation (Ikenberry and Kuh 2015), wherein assessment activities for reporting purposes are decoupled from institutional improvement processes. One innovative way to address this problem is observed in the Multi-State Collaborative to Advance Student Learning (MSC), a national effort involving thirteen states and 900 faculty members at eighty public two- and four-year institutions that aims to improve student learning through direct engagement with faculty and the assessment of authentic student work (Berrett 2016). The MSC eschews the use of standardized exams in favour of a unique assessment approach termed the Valid Assessment of Learning in Undergraduate Education (VALUE). In this approach, faculty evaluators selected from multiple institutions use proficiency criteria specified through VALUE rubrics to rate students' coursework in such areas as critical thinking, writing, and quantitative reasoning. Performance ratings are then shared with both students and their respective instructors, thereby

6 The significant role of accreditation is partly due to exhortations in the 1998 federal reauthorization of the Higher Education Act of 1965 for accreditors to emphasize the evaluation of student achievement in reviews of program integrity.

providing faculty with feedback that can be used to improve specific courses. Notably, current reliability and validity limitations may preclude the use of VALUE rubrics for summative purposes (see Carnahan 2016), but the MSC does offer a formative approach to student assessment that might profitably complement present accountability models.

Conclusion

Few issues in higher education deserve the descriptor of crisis more firmly than the problem of quality. Whether one consults research using standardized exams or surveys of employers, a significant proportion of college graduates do not appear to be meeting learning expectations. Furthermore, challenges of inclusive excellence will most certainly become evident at the intersection of quality and equity as lower-income students are channelled to institutions of questionable effectiveness (Mettler 2014) or lack the resources to fully participate in the learning process (e.g., Whatley 2017). Among the first ameliorative steps would be the development of a comprehensive system for assessing the current state of educational practice, institutional effectiveness, and degree integrity with the aim of shaping expectations, providing evidence of student learning, and stimulating improvement. Nothing short of the very possibility of a robust economy and vibrant democracy is at stake.

References

AAC&U (Association of American Colleges and Universities). 2007. *College Learning for the New Global Century: A Report from the National Council for Liberal Education and America's Promise.* Washington, DC: Association of American Colleges and Universities. https://www.aacu.org/sites/default/files/files/LEAP/GlobalCentury_final.pdf

Adelman, Cliff. 2006. *The Toolbox Revisited: Paths to Degree Completion from High School Through College.* Washington, DC: US Department of Education.

Adelman, Cliff, Peter Ewell, Paul Galston, and Carol G. Schneider. 2014. *The Degree Qualifications Profile*. Indianapolis: Lumina Foundation. https://www.luminafoundation.org/files/resources/dqp.pdf

Arum, Richard, and Josipa Roska. 2011. *Academically Adrift: Limited Learning on College Campuses*. Chicago: University of Chicago Press.

Arum, Richard, Josipa Roska, and Amanda Cook. 2016. *Improving Quality in American Higher Education: Learning Outcomes and Assessments for the 21st Century*. San Francisco: Jossey-Bass.

Asmussen, John G., and Aaron S. Horn. 2014. *Developmental Education: A Review of Research on Programmatic Reforms.* Minneapolis: Midwestern Higher Education Compact. http://files.eric.ed.gov/fulltext/ED566747.pdf

Barkley, Elizabeth F. 2009. S*tudent Engagement Techniques: A Handbook for College Faculty*. San Francisco: Jossey-Bass.

Barrows, Howard S. 1996. "Problem-based Learning in Medicine and Beyond: A Brief Overview." *New Directions for Teaching and Learning* (68): 3–12.

Berrett, Dan. 2016. "The Next Great Hope for Measuring Learning." *The Chronicle of Higher Education*, 16 October. http://www.chronicle.com/article/The-Next-Great-Hope-for/238075

Biggs, John, and Catherine Tang. 2011. *Teaching for Quality Learning at University: What the Student Does,* 4th ed. Berkshire: Open University Press and McGraw-Hill.

Bowman, Nicholas A. 2010. "Can 1st-year College Students Accurately Report Their Learning and Development?" *American Educational Research Journal* 47 (2): 466–96.

———. 2011. "Validity of College Self-reported Gains at Diverse Institutions." *Educational Researcher* 40 (1): 22–4.

Bray, Gayle Babbitt, Ernest T. Pascarella, and Christopher T. Pierson. 2004. "Postsecondary Education and Some Dimensions of Literacy Development: An Exploration of Longitudinal Evidence." *Reading Research Quarterly* 39 (3): 306–30.

Campbell, Corbin M. 2015. "Serving a Different Master: Assessing College Educational Quality for the Public." In *Higher Education: Handbook of Theory and Research* Vol. 30, 525–79. Springer International Publishing.

Campbell, Corbin M., and Alberto F. Cabrera. 2014. "Making the Mark: Are Grades and Deep Learning Related?" *Research in Higher Education* 55 (5): 494–507.

Carnahan, Julie. 2016. *Multi-State Collaborative to Advance Quality Student Learning—Institutes of Education Sciences Grant Application*. Boulder: State Higher Education Executive Officers.

Celio, Christine I., Joseph Durlak, and Allison Dymnicki. 2011. "A Meta-analysis of the Impact of Service-learning on Students." *Journal of Experiential Education* 34 (2): 164–81.

Desjardins, Richard, William Thorn, Andreas Schleicher, Glenda Quintini, Michele Pellizzari, Viktoria Kis, and Ji Eun Chung. 2013. *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*.

Paris: OECD Publishing. https://www.oecd.org/skills/piaac/Skills%20volume%201%20(eng)--full%20v12--eBook%20(04%2011%202013).pdf

Dochy, Filip, Mien Segers, Piet Van den Bossche, and David Gijbels. 2003. "Effects of Problem-based Learning: A Meta-analysis." *Learning and Instruction* 13 (5): 533–68.

Donhardt, Gary L. 2004. "In Search of the Effects of Academic Achievement in Postgraduation Earnings." *Research in Higher Education* 45 (3), 271–84.

Duckworth, Angela L., Patrick D. Quinn, Donald R. Lynam, Rolf Loeber, and Magda Stouthamer-Loeber. 2011. "Role of Test Motivation in Intelligence Testing." *Proceedings of the National Academy of Sciences* 108 (19): 7716–20.

Eagan, Kevin, Ellen B. Stolzenberg, Jennifer B. Lozano, Melissa C. Aragon, Maria R. Suchard, and Sylvia Hurtado. 2014. *Undergraduate Teaching Faculty: The 2013–2014 HERI Faculty Survey*. Los Angeles: Higher Education Research Institute, UCLA. https://pdfs.semanticscholar.org/dede/d6df13d5834cab7df13701affe339a1af8a3.pdf

Ewell, Peter T. 2008. "Assessment and Accountability in America Today: Background and Context." *New Directions for Institutional Research*, 2008 (S1): 7–17.

———. 2009. *Assessment, Accountability and Improvement: Revisiting the Tension*. Champaign, IL: National Institute for Learning Outcomes Assessment. http://www.learningoutcomeassessment.org/documents/PeterEwell_005.pdf

Ewell, Peter T., and Dennis P. Jones. 1996. *Indicators of "Good Practice" in Undergraduate Education: A Handbook for Development and Implementation*. Boulder, CO: National Center for Higher Education Management. http://files.eric.ed.gov/fulltext/ED403828.pdf

Finney, Sara J., Donna L. Sundre, Matthew S. Swain, and Laura M. Williams. 2016. "The Validity of Value-added Estimates from Low-stakes Testing Contexts: The Impact of Change in Test-taking Motivation and Test Consequences." *Educational Assessment* 21 (1): 60–87.

Francis, Perry C., and Aaron Horn. 2017. "Mental Health Issues and Counseling Services in US Higher Education: An Overview of Recent Research and Recommended Practices." *Higher Education Policy* 30 (2): 263–77.

Gonyea, Robert M., and Angie Miller. 2011. "Clearing the AIR about the Use of Self-reported Gains in Institutional Research." *New Directions for Institutional Research* 2011 (150): 99–111.

Gordon, Jonathan, Joe Ludlum, and Joseph J. Hoey. 2008. "Validating

NSSE against Student Outcomes: Are They Related?" *Research in Higher Education*, 49 (1): 19–39.

Hart Research Associates. 2015. *Falling Short? College Learning and Career Success.* Washington, DC: Association of American Colleges and Universities. https://www.aacu.org/sites/default/files/files/LEAP/2015employerstudentsurvey.pdf

Harvey, Lee, and Diana Green. 1993. "Defining Quality." *Assessment & Evaluation in Higher Education* 18 (1): 9–34.

Hillman, Nicholas W., Alisa Hicklin Fryar, and Valerie Crespín-Trujillo. 2017. "Evaluating the Impact of Performance Funding in Ohio and Tennessee." *American Educational Research Journal*. doi:0002831217732951

Horn, Aaron S. 2012. "The Cultivation of a Prosocial Value Orientation through Community Service: An Examination of Organizational Context, Social Facilitation, and Duration. *Journal of Youth and Adolescence* 41 (7): 948–68.

Horn, Aaron S. and Gerald W. Fry. 2013. "Promoting Global Citizenship Through Study Abroad: The Influence of Program Destination, Type, and Duration on the Propensity for Development Volunteerism." *VOLUNTAS* 24 (4), 1159–79.

Horn, Aaron S., Olena G. Horner, and Giljae Lee. 2017. "Measuring the Effectiveness of Two-year Colleges: A Comparison of Raw and Value-added Performance Indicators." *Studies in Higher Education:* 1–19.

Horn, Aaron S., and Giljae Lee. 2016. "The Reliability and Validity of Using Regression Residuals to Measure Institutional Effectiveness in Promoting Degree Completion." *Research in Higher Education* 57 (4): 469–96.

———. 2017. "Evaluating the Accuracy of Productivity Indicators in Performance Funding Models." *Educational Policy.* doi:0895904817719521

Huber, Christopher R., and Nathan R. Kuncel. 2016. "Does College Teach Critical Thinking? A Meta-analysis." *Review of Educational Research* 86 (2): 431–68.

Hyytinen, Heidi, Kari Nissinen, Jani Ursin, Auli Toom, and Sari Lindblom-Ylänne. 2015. "Problematising the Equivalence of the Test Results of Performance-based Critical Thinking Tests for Undergraduate Students." *Studies in Educational Evaluation* 44: 1–8.

Ikenberry, Stanley O., and George D. Kuh. 2015. "From Compliance to Ownership: Why and How Colleges and Universities Assess Student Learning." In *Using Evidence of Student Learning to Improve Higher Education,* edited by George Kuh, Stanley O. Ikenberry, Natasha Jankowski, Timothy Reese Cain, Peter T. Ewell, Pat Hutchings, and

Jillian Kinzie, 1–23. San Francisco: Jossey-Bass.
Inkelas, Karen K., and Matthew Soldner. 2011. "Undergraduate Living-learning Programs and Student Outcomes." In *Higher Education: Handbook of Theory and Research,* edited by John C. Smart and Michael B. Paulsen, 1–55. Springer.
Jankowski, Natasha. 2017. *Unpacking Relationships: Instruction and Student Outcomes.* Washington DC: American Council on Education. http://www.acenet.edu/news-room/Documents/Unpacking-Relationships-Instruction-and-Student-Outcomes.pdf
Johnson, David W., Roger T. Johnson, and Karl A. Smith. 1998. "Cooperative Learning Returns to College; What Evidence Is There That It Works?" *Change* 30 (4): 26–35.
Kelchen, Robert, and Luke J. Stedrak. 2016. "Does Performance-based Funding Affect Colleges' Financial Priorities?" *Journal of Education Finance* 41 (3): 302–21.
Khator, Renu. 2011. "Envisioning America's Global University." *U.S. News and World Report,* 13 September. https://www.usnews.com/education/articles/2011/09/13/envisioning-americas-global-university
Klein, Stephen, Ou L. Liu, and James Sconing. 2009. *Test Validity Study Report*. https://cp-files.s3.amazonaws.com/26/TVSReport_Final.pdf
Kornhauser, Zachary G., Jillian Minahan, Karen L. Siedlecki, and Jeffrey T. Steedle. 2014. *A Strategy for Increasing Student Motivation on Low-stakes Assessments*. Council for Aid to Education. http://cae.org/images/uploads/pdf/A_Strategy_for_Increasing_Student_Motivation.pdf
Kuh, George. D. 2008. *High-impact Educational Practices: What They Are, Who Has Access to Them, and Why They Matter*. Washington, DC: Association of American Colleges and Universities.
Kuh, George. D., Natasha Jankowski, Stanley. O. Ikenberry, and Jillian L. Kinzie. 2014. *Knowing What Students Know and Can Do: The Current State of Student Learning Outcomes Assessment in US Colleges and Universities.* Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).
Kuh, George. D., Jillian Kinzie, John H. Schuh, and Elizabeth J. Whitt. (2011). *Student Success in College: Creating Conditions that Matter*. Hoboken, NJ: John Wiley & Sons.
Kutner, Mark, Elizabeth Greenberg, Ying Jin, Bridget Boyle, Yung-chen Hsu, and Eric Dunleavy. 2007. *Literacy in Everyday Life: Results from the 2003 National Assessment of Adult Literacy (NCES 2007-480)*. US Department of Education. Washington, DC: National Center for Education Statistics. http://files.eric.ed.gov/fulltext/ED495996.pdf

Kyndt, Eva, Elisabeth Raes, Bart Lismont, Fran Timmers, Eduardo Cascallar, and Filip Dochy. 2013. "A Meta-analysis of the Effects of Face-to-face Cooperative Learning. Do Recent Studies Falsify or Verify Earlier Findings?" *Educational Research Review* 10: 133–49.

Levine, Peter. 2007. *The Future of Democracy: Developing the Next Generation of American Citizens*. Medford: Tufts University Press.

Liu, Ou L., Brent Bridgeman, and Rachel M. Adler. 2012. "Measuring Learning Outcomes in Higher Education: Motivation Matters." *Educational Researcher* 41 (9): 352–62.

Liu, Ou L., Joseph A. Rios, and Victor Borden. 2015. "The Effects of Motivational Instruction on College Students' Performance on Low-stakes Assessment." *Educational Assessment* 20 (2): 79–94.

Loes, Chad N., Mark H. Salisbury and Ernest T. Pascarella. 2015. "Student Perceptions of Effective Instruction and the Development of Critical Thinking: A Replication and Extension." *Higher Education* 69 (5): 823–38.

Lumina Foundation. 2016. *States with Higher Education Attainment Goals*. http://strategylabs.luminafoundation.org/wp-content/uploads/2013/10/State-Attainment-Goals.pdf

Markle, Ross, Meghan Brenneman, Teresa Jackson, Jeremy Burrus, and Steven Robbins. 2013. Synthesizing Frameworks of Higher Education Student Learning Outcomes. *ETS Research Report Series* 2013 (2).

Mayer, Richard E. 2011. *Applying the Science of Learning*. Upper Saddle River, NJ: Pearson.

Mayhew, Matthew. J., Nicholas A. Bowman, Alyssa N. Rockenbach, Tricia A. Seifert, and Gregory C. Wolniak. 2016. *How College Affects Students: 21st Century Evidence that Higher Education Works* Vol. 3. San Francisco: Jossey-Bass.

McDonnell, Simon T., Colin C. Chellman, Cheryl B. Littman, and David B. Crook. 2013. Implementing Value-added Accountability Measures at the City University of New York. https://www.cuny.edu/about/administration/offices/ira/opr/papers/accountability_measures.pdf

McCormick Alexander C., Jillian Kinzie, and Robert M. Gonyea. 2013. "Student Engagement: Bridging Research and Practice to Improve the Quality of Undergraduate Education." In *Higher Education: Handbook of Theory and Research* Vol. 28, edited by M. Paulsen. 47–92. Springer: Dordrecht.

Mettler, S. 2014. *Degrees of Inequality: How the Politics of Higher Education Sabotaged the American Dream*. New York: Basic Books.

Miller, Margaret A., and Peter T. Ewell. 2005. *Measuring Up on Col-*

lege-level Learning. National Center for Public Policy and Higher Education. http://www.highereducation.org/reports/mu_learning/Learning.pdf

Minnesota Office of Higher Education. 2012. *Minnesota Measures: 2012 Report on Higher Education Performance*. http://www.ohe.state.mn.us/pdf/MinnesotaMeasures2012.pdf

National Research Council. 2013. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. National Academies Press.

Nelson Laird, Thomas F., Tricia A. Seifert, Ernest T. Pascarella, Matthew J. Mayhew, and Charles F. Blaich. 2014. "Deeply Affecting First-year Students' Thinking: Deep Approaches to Learning and Three Dimensions of Cognitive Development." *The Journal of Higher Education* 85 (3): 402–32.

NSSE (National Survey of Student Engagement). 2017. *NSSE topical modules*. http://nsse.indiana.edu/html/modules.cfm

O'Neil, Harold F., Jamal Abedi, Judy Miyoshi and Ann Mastergeorge. 2005. "Monetary Incentives for Low-stakes Tests." *Educational Assessment* 10 (3): 185–208.

Pascarella, Ernest. T., Mark H. Salisbury, and Charles Blaich. 2011. "Exposure to Effective Instruction and College Student Persistence: A Multi-institutional Replication and Extension." *Journal of College Student Development* 52 (1): 4–19.

Pike, Gary R. 2013. "NSSE Benchmarks and Institutional Outcomes: A Note on the Importance of Considering the Intended Uses of a Measure in Validity Studies." *Research in Higher Education* 54 (2): 149–70.

Porter, Stephen R. 2013. "Self-reported Learning Gains: A Theory and Test of College Student Survey Response." *Research in Higher Education* 54 (2): 201–26.

Reddy, Peter and Elisabeth Moores. 2012. "Placement Year Academic Benefit Revisited: Effects of Demographics, Prior Achievement and Degree Programme." *Teaching in Higher Education* 17 (2): 153–65.

Reindl, Travis, and R. Reyna. 2011. Complete to Compete: From Information to Action—Revamping Higher Education Accountability Systems. *NGA Center for Best Practices*.

Rios, Joseph A., Ou L. Liu, and Brent Bridgeman. 2014. "Identifying Low-Effort Examinees on Student Learning Outcomes Assessment: A Comparison of Two Approaches." *New Directions for Institutional Research* 2014 (161): 69–82.

Roohr, Katrina C., Huili Liu, and Ou L. Liu. 2016. "Investigating Student Learning Gains in College: A Longitudinal Study." *Studies in*

Higher Education 2016: 1–17.

Rutherford, Amanda, and Thomas Rabovsky. 2014. "Evaluating Impacts of Performance Funding Policies on Student Outcomes in Higher Education." *The ANNALS of the American Academy of Political and Social Science* 655 (1): 185–208.

Schneider, Mark, and Lu Yin. 2011. The High Cost of Low Graduation Rates: How Much Does Dropping Out of College Really Cost? *American Institutes for Research*.

Shanghai Ranking Consultancy. 2016. *Academic Ranking of World Universities*. http://www.shanghairanking.com/#

SHEEO (State Higher Education Executive Officers Association). 2005. Accountability for Better Results: A National Imperative for Higher Education. http://www.sheeo.org/resources/publications/accountability-better-results

Sitzmann, Traci, Katherine Ely, Kenneth G. Brown, and Kristina N. Bauer. 2010. "Self-assessment of Knowledge: A Cognitive Learning or Affective Measure?" *Academy of Management Learning & Education* 9 (2): 169–91.

Sparks, Erin and Mary Jo Waits. 2011. Degrees for What Jobs? Raising Expectations for Universities and Colleges in a Global Economy. *NGA Center for Best Practices*.

Swerdzewski, Peter. J., J. Christine Harmes, and Sara J. Finney. 2011. "Two Approaches for Identifying Low-motivated Students in a Low-Stakes Assessment Context." *Applied Measurement in Education* 24 (2): 162–88.

Tandberg, David A., and Nicholas W. Hillman. 2014. "State Higher Education Performance Funding: Data, Outcomes, and Policy Implications." *Journal of Education Finance* 39 (3): 222–43.

THEC (Tennessee Higher Education Commission). 2015. *Quality Assurance Funding: 2015–20 Cycle Standards.* https://www.tn.gov/assets/entities/thec/attachments/THEC_2015-20_Quality_Assurance_Funding_Guidebook.pdf

Tinto, Vincent. 2012. *Completing College: Rethinking Institutional Action*. Chicago: University of Chicago Press.

US Department of Education. 2006. *A Test of Leadership: Charting the Future of US Higher Education.* https://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf

Umbricht, Mark R., Frank Fernandez, and Justin C. Ortagus. 2017. "An Examination of the (Un) Intended Consequences of Performance Funding in Higher Education." *Educational Policy 31* (5): 643–73.

Warren, Jami L. 2012. "Does Service Learning Increase Student Learning?: A Meta-analysis." *Michigan Journal of Community Service Learning* 18 (2): 56–61.

Weiss, Michael J., Alexander K. Mayer, Dan Cullinan, Alyssa Ratledge, Colleen Sommo, and John Diamond. 2015. "A Random Assignment Evaluation of Learning Communities at Kingsborough Community College—Seven Years Later." *Journal of Research on Educational Effectiveness* 8 (2): 189–217.

Whatley, Melissa 2017. "Financing Study Abroad: An Exploration of the Influence of Financial Factors on Student Study Abroad Patterns." *Journal of Studies in International Education*. doi:1028315317697798

Wise, Steven L., and Christine E. DeMars. 2005. "Low Examinee Effort in Low-stakes Assessment: Problem and Potential Solutions." *Educational Assessment* 10 (1): 1–17.

Yorio, Patrick L., and Feifei Ye. 2012. "A Meta-analysis on the Effects of Service Learning on the Social, Personal, and Cognitive Outcomes of Learning." *Academy of Management Learning & Education* 11 (1): 9–27.

Zhang, Liang. 2008. "The Way to Wealth and the Way to Leisure: The Impact of College Education on Graduates' Earnings and Hours of Work." *Research in Higher Education* 49 (3): 199–213.

Zilberberg, Anna, Sara J. Finney, Kimbery R. Marsh, and Robin D. Anderson. 2014. "The Role of Students' Attitudes and Test-taking Motivation on the Validity of College Institutional Accountability Tests: A Path Analytic Model." *International Journal of Testing* 14 (4): 360–84.

Zilvinskis, John, Anthony A. Masseria, and Gary R. Pike. 2017. "Student Engagement and Student Learning: Examining the Convergent and Discriminant Validity of the Revised National Survey of Student Engagement." *Research in Higher Education* 2017: 1–24.

Zukin, Cliff, Scott Keeter, Molly Andolina, Krista Jenkins, and Michael X. Delli Carpini. 2006. *A New Engagement? Political Participation, Civic Life, and the Changing American Citizen*. Oxford University Press.