

DRAFT: Please do not cite without permission of the authors.

Measuring College Performance

Richard Arum  
New York University

Josipa Roksa  
University of Virginia

November 2011

\*Paper prepared for “Mapping Broad-Access Higher Education” conference at Stanford University, December 1-2, 2011.

## Measuring College Performance

Public and private investment in higher education, the significance of higher educational attainment for individual life course outcomes and the presumed role of human capital formation for economic competitiveness has led to increasing interest in measuring college quality. It is thus not surprising that while metrics of institutional quality in higher education have existed and have shaped organizational orientations and practices for decades, something has changed in the organizational environment facing colleges and universities.

Today, there is a growing discussion about moving towards greater organizational accountability in higher education. For example, the Spellings Commission report *A Test of Leadership: Charting the Future of Higher Education* in 2006 noted “that students, parents, and policymakers are often left scratching their heads over the answers to basic questions” given the “lack of clear, reliable information about the cost and quality of postsecondary institutions, along with a remarkable absence of accountability mechanisms to ensure that colleges succeed in educating students.”<sup>1</sup> Since the Commission’s report, various external assessments and accountability measures have been proposed or advanced at all levels of the system.

For example, The Association of Public and Land-grant Universities (APLU) and the Association of State Colleges and Universities (AASCU) in 2007 launched a Voluntary System of Accountability (VSA) that led several hundred four-year colleges and universities to assess and make public information on standardized student learning

---

<sup>1</sup> U.S. Department of Education, *A Test of Leadership: Charting the Future of U.S. Higher Education* (Washington, DC: U.S. Department of Education, 2006), x.

assessments that were administered at their institutions. Public colleges and universities in Texas have been required to publicly report course syllabi, faculty research grants and teaching loads as well as course evaluations – the latter of which were used to measure and financially reward faculty members’ assumed commitment to teaching. The federal government has also continued to move in this direction with the recent adoption of rules that selectively limit federal student loan eligibility to for-profit and other vocational programs – requiring that these institutions demonstrate gainful employment of graduates (measured by whether 35 percent of graduates three years out are repaying student loan principal and whether the average graduate’s loan payment does not exceed 30 percent of discretionary income or 12 percent of total earnings).<sup>2</sup>

In multiple public forums, we have vehemently argued against the desirability of externally imposed accountability schema. We are deeply skeptical of increased centralized regulation of this character – fearing that the unintended consequences would far outweigh any benefits – and have instead called for accountability to be enhanced at the institutional level. Moreover, most of the institutional measures previously used to measure college quality were generally not designed for accountability purposes. Nevertheless, given current policy attention, we discuss various dimensions of assessment and accountability in this paper to contribute to the broader policy discourse and public understanding of the subject.

We believe that at this historic juncture, social scientists could play a useful public role by describing existing conditions in higher education as well as the organizational incentives associated with observed learning outcomes. In addition, social

---

<sup>2</sup> Tamar Lewin, “Education Department Increases Its Regulation of For-Profit Colleges,” *New York Times* (June 2, 2011).

scientists can contribute to policy formation by identifying the strengths and limitations of various metrics currently employed or under consideration with respect to their underlying assumptions, measurement properties and ability to shape organizational behavior.

*Contemporary Conditions and Organizational Incentives*

As accountability pressures are mounting, so are reports of limited learning in higher education. *A Test of Leadership*, a report by the Secretary of Education's Commission on the Future of Higher Education, was one of the most visible indictments of higher education, placing the issue of learning on the national agenda. Using data from the National Assessment of Adult Literacy, the commission argued that "the quality of student learning at U.S. colleges and universities is inadequate, and in some cases, declining."<sup>3</sup> In the same year, the former President of Harvard University Derek Bok, lamented in an aptly titled book *Our Underachieving Colleges* that many students today are graduating from college "without being able to write well enough to satisfy their employers...reason clearly or perform competently in analyzing complex, non-technical problems."<sup>4</sup> Recent studies of collegiate learning provide further evidence for this claim.

In *Academically Adrift*, Richard Arum and Josipa Roksa, followed over 2,000 students through a diverse set of four-year institutions and found that students demonstrated limited or no growth on the Collegiate Learning Assessment (CLA), an objective measure of critical thinking, complex reasoning and written communication,

---

<sup>3</sup> U.S. Department of Education, *A Test of Leadership: Charting the Future of U.S. Higher Education* (Washington, DC: U.S. Department of Education, 2006), 3.

<sup>4</sup> Derek Bok, *Our Underachieving Colleges: A Candid Look at How Much Students Learn and Why They Should Be Learning More* (Princeton, NJ: Princeton University Press, 2006), 8.

during their first two years of college.<sup>5</sup> Subsequent data analyses considering all four years of college reveal similarly disturbing patterns: thirty-six percent of students do not demonstrate any significant improvement in learning, as measured by CLA performance, over the whole four years of college.<sup>6</sup> Moreover, a recent replication study, using data from the Wabash National Study of Liberal Arts Education and relying on a different measure of learning – the Collegiate Assessment of Academic Proficiency Critical Thinking Test – reports equally limited improvement in critical thinking among students attending four-year institutions.<sup>7</sup> The consistency of findings across datasets and measures eliminates the likelihood that the results are produced by unique properties of assessment tools or samples studied. Instead, the evidence points to what many have suspected but have not been able to demonstrate, that something is amiss in American higher education.

While this pattern of limited learning may be disturbing, it is not necessarily surprising considering current incentive structures in higher education. College students today invest only modest levels of effort in their studies. Sophomores in *Academically Adrift* reported studying on average only 12 hours per week, one third of which was spent studying with peers. Even more alarming, 37 percent dedicated five or fewer hours per week to studying alone. This level of limited study time persists through the senior year and is consistent with other reports of students' time use, such as those observed in the *National Survey of Student Engagement* (NSSE). Similarly, a recent study of the

---

<sup>5</sup> Richard Arum and Josipa Roksa, *Academically Adrift: Limited Learning on College Campuses* (Chicago: University of Chicago Press, 2011).

<sup>6</sup> Richard Arum, Josipa Roksa, and Esther Cho. 2011. *Improving Undergraduate Learning: Findings and Policy Recommendations from the SSRC-CLA Longitudinal Project*. New York: Social Science Research Council.

<sup>7</sup> Ernest Pascarella et al. 2011. "How Robust are Findings of Academically Adrift?" *Change Magazine* May/June: 20-24.

University of California undergraduates reported that while students spent thirteen hours a week studying, they also spent twelve hours socializing with friends, eleven hours using computers for fun, six hours watching television, six hours exercising, five hours on hobbies and three hours on other forms of entertainment. Students were thus spending on average 43 hours per week outside of the classroom on these activities – i.e., over three times more hours than the time they spent studying.<sup>8</sup>

Students can dedicate modest time to their studies in part because they are able to find courses and programs that place limited demands on them. Fifty percent of sophomores in *Academically Adrift* reported that they had not taken a single course the prior semester that required more than twenty pages of writing over the course of the semester; one-third did not take a single course the prior semester that required on average even more than 40 pages of reading per week. Over their entire four years of college coursework, 50 percent of students reported that they had taken five or fewer courses that required 20 pages of writing over the course of the semester, and 20 percent of students reported that they had taken five or fewer courses that required 40 pages of reading per week. These patterns of meager academic requirements are found in other national studies, such as the *National Survey of Student Engagement*.

Students are able to navigate the higher education system today without experiencing substantial academic demands or investing long hours in their studies. And moreover, they are not only able to graduate but are also able to attain a high GPA in the process. Among students in *Academically Adrift* who spent five or fewer hours studying alone, the average GPA was 3.2. These patterns can be understood in part as a reflection

---

<sup>8</sup> Steven Brint and Allison M. Cantwell, “Undergraduate Time Use and Academic Outcomes: Results from UCUES 2006.” Research and Occasional Paper Series (Center for Students in Higher Education, University of California, Berkeley, 2008).

of what George Kuh has termed a “disengagement compact” – an implicit agreement between faculty and students to minimize academic demands: “‘I’ll leave you alone if you leave me alone.’ That is, I won’t make you work too hard (read a lot, write a lot) so that I won’t have to grade as many papers or explain why you are not performing well.”<sup>9</sup>

While it can be argued that American higher education has always had a strong social component which detracted from academics, recent decades have seen shifts in several key indicators suggesting that the current context is indeed different from the past. Labor economists Philip Babcock and Mindy Marks, for example, have conducted meticulous analyses of student time-use data from twelve individual-level surveys from the 1920s to today. They have found that full-time college students through the early 1960s spent roughly 40 hours per week on academic pursuits (i.e., combined studying and class time), at which point a steady decline ensued throughout the following decades. Average time studying fell from 25 hours per week in 1961 to 20 hours per week in 1981 and 13 hours per week in 2003.<sup>10</sup> At the same time that studying has been decreasing, student expectations have been on the rise. Higher Education Research Institute (HERI) at UCLA has been asking freshmen during orientation week about the grades they expect to get in college. In 1971, when the question was first asked, only 27 percent of students expected to have at least a B average in college. Two thirds of students have that expectation today.<sup>11</sup> And they don’t get disappointed. Despite limited time investment in academic pursuits, grade inflation has been on the rise, with the vast majority of grades today being Bs or higher. A recent study tracking grade inflation over time indicates that

---

<sup>9</sup> George Kuh, “What We Are Learning About Student Engagement,” *Change* 35 (2003): 28.

<sup>10</sup> Philip Babcock and Mindy Marks, “The Falling Time Cost of College: Evidence from Half a Century of Time Use Data,” *Review of Economics and Statistics* (forthcoming).

<sup>11</sup> Higher Education Research Institute (HERI). (2009). *The American college teacher: National norms for 2007–2008*. HERI, University of California at Los Angeles: Los Angeles, CA.

while only 15 percent of all letter grades in the 1960s were A's, today A's account for 43 percent of grades. In the meantime, the proportion of B's has stayed stable and that of C's has declined precipitously to accommodate the rise of A's. The study reported that approximately three-quarters of letter grades today are above a C – and only a few decades ago, C used to imply “average.”<sup>12</sup> If grades were perceived solely as a positional good, this would not be a problem. Students, however, often engage in satisficing –with a B grade, attainable with little effort, considered sufficient.<sup>13</sup>

Limited learning on college campuses is a product of many forces, but one deserving a particularly close scrutiny is a change in the relationship between students and institutions over the course of the 20<sup>th</sup> century. Students in higher education are increasingly defined as “consumers” and “clients.” In this context, schools are not expected to provide quasi-parental guidance and social regulation, but instead to meet client needs through delivery of elaborate and ever expanding services. Accordingly, colleges and universities have increasingly diverted resources toward non-academic functions. Gary Rhoades has documented that over the past three decades non-faculty support professionals have become the fastest growing category of professional employment in higher education, with the most significant increase occurring in the broad area of student services, including admissions, financial aid, career placement, counseling and academic services (such as advising and tutoring that have been reassigned to non-faculty professionals).<sup>14</sup>

---

<sup>12</sup> Stuart Rojstaczer and Christopher Healy. 2012. “Where A Is Ordinary: The Evolution of American College and University Grading, 1940-2009.” *Teachers College Record* 114: TBD [currently published on line]

<sup>13</sup> Phillip Babcock, “Real Costs of Nominal Grade Inflation? New Evidence from Student Course Evaluations,” *Economic Inquiry* (2010) 48/4.

<sup>14</sup> Gary Rhoades, “The Study of American Professions,” in *Sociology of Higher Education: Contributions and their Contexts*, ed. Patricia Gumpert (Baltimore, MD: Johns Hopkins Press, 2007).

Moreover, universities have increasingly turned toward part-time instructors for teaching. The percentage of part-time faculty is nearly half of all faculty and instructional staff in higher education today.<sup>15</sup> One of the consequences of this shift is an increasing receptiveness of faculty to the whims of students as consumers. The livelihood of part-time faculty is highly dependent on students' course evaluations, which are often used as the only (or most important) metric of teaching quality. If students as consumers emphasized rigorous instruction, this could have positive consequences. However, a market-based logic of education encourages students to focus on the instrumental value of education – i.e., obtaining credentials – and doing so with the lowest possible investment of time and energy.<sup>16</sup> Consequently, students' enrollment decisions are related to the leniency of the course, and their course evaluations are related to the grades they believe they will receive at the end.<sup>17</sup> Faculty, particularly those whose lives are more closely tied to students' course evaluations, thus have an incentive to offer easy classes accompanied by good grades, creating a downward spiral in academic rigor.

While part-time faculty's lives are closely tied to student satisfaction, full-time faculty's lives are increasingly related to their research productivity. Although an expected emphasis at research institutions, Ernest Boyer's work in the late 1980s highlighted the changing "priorities of the professoriate" as well as the institutional diffusion of the university research model to faculty at institutions throughout the system. An increasing number of faculty across four-year institutions reported that scholarship was more important than teaching for tenure decisions in their departments. Other

---

<sup>15</sup> U.S. Department of Education. 2009. *Digest of Education Statistics*. (Washington DC: U.S. Department of Education).

<sup>16</sup> David Labaree, *How to Succeed in School without Really Learning: The Credentials Race in American Education* (New Haven, CT: Yale University Press, 1997).

<sup>17</sup> Valen E. Johnson, *Grade Inflation: A Crisis in College Education* (New York: Springer-Verlag, 2003).

developments in higher education, including the commercialization of higher education following the Bayh-Dole Act of 1980<sup>18</sup> and the “output creep” through which faculty have gained increased discretionary time to pursue their professional goals,<sup>19</sup> have all combined to focus faculty on activities not related to undergraduate education.

Not all institutions are equally susceptible to these pressures and not all of them have responded in the same way or to the same extent. However, isomorphic pressures in the field of higher education have compelled most institutions to move in similar directions of treating students as consumers, divesting from academic functions, and increasing demands on faculty that are not related to undergraduate learning. Institutions that are likely most susceptible to these pressure are open-access colleges. While prestigious institutions reject 60-90 percent of the students who apply, open-access colleges depend on enrollments. If students as consumers expect certain luxuries and services, combined with a credential for limited efforts, these institutions are in the weakest position to resist. Decreases in state and federal funding in recent decades have only increased their reliance on tuition. Moreover, faced with limited resources, open-access institutions have often turned to large classes, many of which are taught by part-time instructors. At the same time, students they enroll require more, not less resources focused on academics, given the often inadequate preparation they receive in K-12 education. Open-access institutions are thus in a particularly precarious position that makes them especially susceptible to recent trends in higher education.

---

<sup>18</sup> Walter Powell and Jason Owen-Smith, “The New World of Knowledge Production in the Life Sciences,” in *The Future of the City of Intellect: The Changing American University*, ed. Steven Brint (Stanford, CA: Stanford University Press, 2002), 115.

<sup>19</sup> William F. Massy and Robert Zemsky, “Faculty Discretionary Time: Departments and the Academic Ratchet,” *Journal of Higher Education* 65 (1994): 1-22.

Organizational changes in higher education have also had unique consequences for administrators, who have been distracted with multiple demands, few of which are related to the quality of undergraduate education. Notably, administrators have a responsibility for the overall stature of their institutions – and therefore how status and recognition are awarded has consequences for their actions. If a particular set of characteristics are regarded as desirable (i.e., if they place institutions higher in public rankings such as U.S. World and News and satisfy external demands from legislators and other interested parties), administrators will have an incentive to work on improving those outcomes. Specific definitions of college quality thus provide incentives for administrators to orient institutional action in particular directions. To date, however, external incentives focused on “college quality” have detracted from a focus on undergraduate learning.

### *Normative Dimensions of Assessments*

Identification and utilization of college quality metrics must always inherently rely on normative assumptions – albeit often unstated – about what the purposes and functions of higher education should be. Resistance to assessment measures is thus often the result of underlying disagreements about the desirable aims of higher education, rather than objections to the technical character of the measurements *per se*. We believe it is useful to highlight, rather than leave unstated the normative dimension of assessment and accountability efforts.

Social scientists and historians have highlighted the extent to which the education sector as a whole has been characterized by difficulties in defining explicitly agreed upon

and measurable outcomes. “The history of higher education,” according to David Labaree, “has been a tale of ambivalent goals and muddled outcomes.”<sup>20</sup> Given this ambiguity, education systems have developed organizational practices that are “loosely coupled,” where institutional legitimacy is gained not by measurement of efficient performance, but through “ceremonial rituals” and adoption of taken-for-granted “isomorphic practices.”<sup>21</sup>

It is also worth emphasizing here, as we have elsewhere, that one would be wrong to assume that higher education is primarily focused on outcomes associated with enhancing undergraduate student learning. The higher education system as a whole has multiple functions including generating scientific discoveries – potentially contributing to economic development – as well as producing general knowledge and local forms of cultural enrichment and entertainment (including, of course, athletic spectacles). Mitchell Stevens and his colleagues have highlighted “the plurality of institutional domains in which higher education is implicated” noting that higher education systems can be conceived of as hubs “connecting multiple social processes that often are regarded as distinct” as well as “sieves for regulating the mobility processes underlying the allocation of privileged positions in the society, incubators for the development of competent social actors, and temples for the legitimation of official knowledge.”<sup>22</sup>

---

<sup>20</sup> David Labaree. 1997. “Public Goods, Private Goods: The American Struggle Over Educational Goals,” *American Educational Research Journal* 34/1:39-81; p. 41.

<sup>21</sup> John Meyer and Brian Rowan. 1977. “Institutionalized Organizations: Formal Structure as Myth and Ceremony,” in *Organizational Environments: Ritual and Rationality*, edited by J. Meyer and R. Scott. Beverly Hills: Sage; pp. 21-44. Paul DiMaggio and Walter Powell. 1983. “The Iron Cage Revisited: Institutional Isomorphism And Collective Rationality In Organizational Fields,” *American Sociological Review* 48:147-60. Karl Weick. 1976. “Educational Organizations as Loosely Coupled Systems,” *Administrative Science Quarterly* 21:1-19.

<sup>22</sup> Stevens, Armstrong and Arum 2008:128.

Focusing solely on the subset of higher education goals related to the development of students, David Labaree has suggested that expectations of the normative functions of U.S. schools have varied historically. Specifically, Labaree highlights the extent to which, while educational objectives have always been contested, overall the emphasis has shifted from the state’s interest in the preparation of democratic citizens, to taxpayers interest in efficiently producing graduates for a stratified occupational structure to a consumer orientation where students aspire solely to gain credentials that are useful for individual social mobility. According to Labaree, this latter orientation can lead to a type of “credentialism that is strikingly counterproductive for both education and society.”<sup>23</sup>

Labaree’s conceptualization of historic variation in the normative orientation of higher education, while helpful and informative, under-emphasizes the extent to which the moral development of students – not just their civic development – was historically a primary organizational goal. The explicitly moral dimensions of these institutions have been most clearly highlighted in Julie Reuben’s work. Reuben notes that prior to World War II, “university reformers continued to view piety and moral discipline as one of the aims of higher education, but wanted to replace older, authoritarian methods with new ones.”<sup>24</sup> Universities saw as their mission helping “people live properly” and attempted to achieve these ends through curricular as well as extra-curricular avenues.<sup>25</sup>

Recent policy initiatives have attempted to move past prior contested, contradictory or ambiguous goals by working to define and articulate a normative

---

<sup>23</sup> Labaree (1997); p. 73.

<sup>24</sup> Julie A. Reuben, *The Making of the Modern University: Intellectual Transformation and the Marginalization of Morality* (Chicago: University of Chicago Press; 1996); p. 12.

<sup>25</sup> Reuben (1996); p. 8.

consensus around student learning outcomes. The Association of American Colleges and Universities (AAC&U), for example, engaged a broad set of institutions over multiple years to advance “a set of educational outcomes that all students need from higher learning” with competencies “keyed to work, life, and citizenship.” The LEAP (Liberal Education and America’s Promise) Initiative in particular advanced the following four broad domains as its “essential learning outcomes”: knowledge of human capital and the physical and natural world; intellectual and practical skills; personal and social responsibility; and integrative and applied learning. While the AAC&U on its website notes that the learning objectives defined in this LEAP Initiative “are best developed by a contemporary liberal education” in their formal report the organization asserts that “the recommended learning outcomes can and should be achieved through many different programs of study and in all collegiate institutions, including colleges, community colleges and technical institutes, and universities, both public and private.”<sup>26</sup>

In addition to historic variation in normative goals, of course, there is also significant institutional variation in higher education. The U.S. higher education system in a comparative framework is particularly noted for its institutional differentiation with schools often focused on distinct types of goals for their students. The Lumina Degree Qualifications Profile attempts to build on the earlier LEAP Initiative to advance a normative framework for higher education that is coherent and shared, but also recognizes such distinctions through highlighting variation in competencies that should

---

<sup>26</sup> American Association of Colleges and Universities, *College Learning for the New Global Century* (Washington, DC: American Association of Colleges and Universities; 2007), p. 2. And <http://www.aacu.org/leap/vision.cfm> (accessed on 11/10/11).

be assessed by different types of institutions.<sup>27</sup> The broad domains for student learning, however, share much in common with the earlier LEAP Initiative with the following categories employed: specialized knowledge; broad, integrative knowledge; intellectual skills; applied learning; civic learning; and an undefined “institution specific areas” component.<sup>28</sup>

Setting aside the Lumina, AAC&U and other efforts to forge a normative consensus around student learning outcomes, the current era is arguably distinct in another important way: today’s discourse on higher education is impoverished for a diverse set of causes. First, the dominance of a neo-liberal market based logic has marginalized practitioners’ concerns that empowering students as consumers does not always lead to educational outcomes that are aligned with broader normative expectations. Second, the end of the Cold War has eliminated an underlying rationale for political and state support of these institutions. Third, possibly as a product of the organizational maturity of the sector, institutional leadership has typically been professionalized and bureaucratized. In spite of a few notable exceptions (e.g., Derek Bok, Michael Crow, Carol Geary Schneider, etc.), the changes in the functions of administrative leadership have led to a decline in the prominence of figures (such as Robert Hutchins or Clark Kerr) that saw their roles as including responsibility for defining and articulating an organizational vision for higher education.<sup>29</sup> The currently

---

<sup>27</sup> Richard Arum, Adam Gamoran and Yossi Shavit. “Inclusion and Diversion in Higher Education: A Study of Expansion and Stratification in 15 Countries.” pp. 1-35 in Shavit, Yossi, Richard Arum and Adam Gamoran, eds. 2007 [2010 paperback]. *Stratification in Higher Education: A Comparative Study*. Palo Alto: Stanford University Press.

<sup>28</sup> Lumina Foundation, *The Degree Qualifications Profile* (Indianapolis, IN: Lumina Foundation, 2011).

<sup>29</sup> David Greenberg, “Small Men on Campus: The Shrinking College President,” *The New Republic* (June 1, 1998).

used measures of college quality have thus often emerged from without as opposed to within higher education.

### *Current Measures of College Quality*

Popularized institutional ranking exercises, such as those undertaken by the U.S. News and World Report, best represent the conception of college quality permeating academic and policy discourses throughout most of the 20<sup>th</sup> century. The metrics utilized by the U.S. News and World Report focus extensively on organizational resources, inputs, and reputation. The input measures have received a disproportionate amount of attention given that many institutions are constrained in their ability to increase substantially their financial resources and reputations are slow to change. Moreover, input measures are most visible, as every year a new cohort of high school graduates enters an anxiety inducing chaos that characterizes access to the institutions at the top of the U.S. Word and News rankings.<sup>30</sup> Input measures highlight characteristics of the incoming students, such as their SAT/ACT scores and high school class standing, as well as institutional acceptance rates. This encourages colleges and universities to focus on increasing the number of applications, rejection rates, yields, and entering students' test scores. And indeed, colleges and universities have paid close attention to these rankings and invested much energy in fashioning their recruitment and admission routines to improve (or at least maintain) their location in the status hierarchy.<sup>31</sup>

---

<sup>30</sup> James Fallows. 2003. "The New College Chaos." *The Atlantic Monthly*, available at: <http://www.theatlantic.com/past/docs/issues/2003/11/fallows.htm>.

<sup>31</sup> Mitchell Stevens. 2007. *Creating a Class: College Admissions and the Education of Elites*. Cambridge, MA: Harvard University Press.

The underlying assumption of the U.S. News and World Report strategy is that abundant resources and highly-selected inputs produce good outcomes. The role of resources in producing desirable educational outcomes, however, has been questioned since the early research on K-12 education in the 1960s. Peers on the other hand represent a significant component of measureable school effects, and some of their characteristics are easily measurable. Focusing on the quality of the student body – although held in disdain by most academics – is thus not without merit. Regardless of whether these inputs actually track with institutional performance, the measures cannot serve as an adequate basis of an accountability framework as they do not measure institutional performance directly nor are they capable of being perceived as legitimate for such purposes. Recruiting a talented pool of students says nothing about what those students have gained and how they have benefited from attending specific institutions.

At the end of the 20<sup>th</sup> century, public and policy attention began to slowly shift from focusing on inputs to asking questions about outcomes, particularly graduation rates. The U.S. Department of Education began requesting information on graduation rates and reporting those publicly in the Graduation Rate Survey. State departments of education similarly began collecting data on graduation rates and policy makers expressed an increasing interest in tying state expenditures on higher education to institutional graduation rates. The U.S. News and World Report responded too: it now includes indicators of the freshman retention rate (the percentage of first-year freshmen who returned to the same college or university the following fall) and overall graduation rate (the percentage of entering freshmen who graduated within six years). These

outcome measures make a sizable contribution to the rankings, as they comprise 20 percent of the total score.<sup>32</sup>

The conceptual shift leading to a focus on outcomes as opposed to inputs is important, but the measurement is problematic. Graduation rates not adjusted for student backgrounds are misleading measures of institutional performance. Student background characteristics, and particularly their academic preparation, account for much of the variation in their likelihood of graduation. While recent research suggests that institutions do have an impact on graduation rates, that impact fades in comparison to the effect of student characteristics.<sup>33</sup> Raw graduation rates are very good proxies of student characteristics and thus inadvertently contribute to an emphasis on inputs. The easiest way to increase graduation rates is to select certain types of students.

Student characteristics can be controlled for, albeit imperfectly, but resulting outcomes are inconsistent and much less easily interpretable. The most recent rendition of the U.S. World and News Report indeed includes a “graduation rate performance” metric for national universities and liberal arts colleges. The graduation rate performance is described as the difference between the actual six-year graduation rate and the predicted graduation rate, which is based upon characteristics of the entering class as well as characteristics of the institution.<sup>34</sup> An obvious question is about what characteristics are included and whether they adequately control for students’ backgrounds. State data systems often have even less information about student backgrounds and thus even less

---

<sup>32</sup> Robert Morse. 2010. Methodology: Undergraduate Ranking Criteria and Weights. U.S. News and World Report, available at: <http://www.usnews.com/articles/education/best-colleges/2010/08/17/methodology-undergraduate-ranking-criteria-and-weights-2011.html>

<sup>33</sup> William Bowen, Matthew Chingos, and Michael McPherson. 2010. *Crossing the Finish Line: Completing College at America’s Public Universities*. Princeton, NJ: Princeton University Press.

<sup>34</sup> *Ibid.*

ability to adjust graduation rates for student characteristics. Moreover, all of the current measures of institutional graduation rates ignore the fact that over half of the students in higher education attend more than one institution. Worse still, reliance on these metrics equates institutional performance with an outcome that implicitly encourages lowering standards and increasing organizational investment in social as opposed to academic functions as social engagement is a strong predictor of student attrition.<sup>35</sup>

When focusing neither on inputs nor outputs provides a compelling indicator of institutional performance, an obvious next step is to consider what happens inside higher education institutions. This is also an empirically compelling shift since most of the variation in student outcomes, from persistence and graduation rates to learning, is within institutions, not across them.<sup>36</sup> Creation of the National Survey of Student Engagement (NSSE) in 2000 drew attention to students' activities within institutions.<sup>37</sup> NSSE was designed to measure student engagement in college and to capture what research suggested were good institutional practices. It was not designed as an accountability measure, but as an institutional tool to provide colleges and universities with information about various student activities and institutional services. Using NSSE to make comparisons across institutions faces some of the same challenges as does using graduation rates – student engagement in particular activities likely has more to do with student characteristics and motivation than institutional practices. Not adjusting responses for student characteristics leaves open the possibility that institutions are

---

<sup>35</sup> Vincent Tinto, *Leaving College: Rethinking the Causes and Cures of Student Attrition* (Chicago: University of Chicago Press, 1993).

<sup>36</sup> Arum and Roksa, *Academically Adrift*. Kevin Carey. 2004. *A Matter of Degrees: Improving Graduation Rates in Four-Year Colleges and Universities*. Washington, D.C.: Education Trust. National Survey of Student Engagement. 2008. *Promoting Engagement for All Students: The Imperative to Look Within*. Bloomington, IN: Center for Postsecondary Research, Indiana University Bloomington.

<sup>37</sup> Nicholas Confessore. 2003. "What Makes A College Good?" *The Atlantic Monthly*, available at: <http://www.theatlantic.com/magazine/archive/2003/11/what-makes-a-college-good/2814/>

getting credit for selecting well, not necessarily contributing to student development above and beyond those initial individual-level inputs.

More problematic from the accountability perspective is that NSSE asked students to self-report their learning during college. Students overall reported that they have learned a substantial amount, and that they have notably improved their higher-order collegiate skills such as critical thinking and analytical reasoning. This has provided false assurance to higher education institutions, suggesting that they do not have to worry about academic rigor or student learning outcomes. Little notice was paid to an obvious contradiction – NSSE survey responses also indicate that students spend a limited amount of time studying and are infrequently asked to do complex, higher-order thinking inside or outside of the classroom.<sup>38</sup>

So how exactly are they learning so much and developing these higher order skills? The NSSE answer seems to be: they are learning outside of the classroom, especially in interactions with their peers. This logic was used to justify increasing spending on social integration and student services. It has thus helped to shift the national attention increasingly away from academic and toward social realms of college. Keeping students engaged became synonymous with keeping students engaged socially, without much regard for academics. Social engagement has its place in college, but its contribution to developing students' academic skills and attitudes is at best questionable. Indeed, recent empirical analyses examining the relationship between a range of NSSE measures and an objective measure of critical thinking have raised questions about the utility of student activities captured by NSSE for improving students' higher-order

---

<sup>38</sup> National Survey of Student Engagement. 2007. *Experiences that Matter: Enhancing Student Learning and Success*. Bloomington, IN: Center for Postsecondary Research, Indiana University Bloomington.

skills.<sup>39</sup> The NSSE experience highlights the challenges and potential pitfalls of relying solely on student self-reports of their activities and outcomes, often examined in a cross-sectional framework, for accountability purposes.

### *New Approaches to Measuring College Quality*

Given increasing scrutiny of higher education and calls for accountability, the question of defining and measuring college quality is crucial. Current indicators, whether considering inputs, graduation rates, or relying on cross-sectional student self-reports, are inadequate for the task. Moreover, the current choice of measures seems a product more of a historical coincidence and political expediency than serious consideration of what higher education should accomplish and how those goals may be assessed. Having a conversation about the purposes of higher education is a difficult task, in part due to the autonomy of individual institutions (and departments and faculty within institutions). It seems clear, however, that an adequate system of accountability will have to take goals of higher education seriously. The key question is thus: what are the goals of higher education, and in particular, what skills, attitudes, and dispositions should be expected from college graduates? Moreover, how well are institutions preparing college graduates in these different realms?

### Graduate Wages

One widely agreed upon goal of higher education (even if contested by academics) is to prepare students for the labor market. This has in recent years led to a

---

<sup>39</sup> Robert M. Carini, George D. Kuh, and Stephen P. Klein. 2006. “Student Engagement and Student Learning: Testing the Linkages.” *Research in Higher Education* 47: 1-32.

proposal to use graduate wage data as a metric of institutional performance. Following the standard human capital framework, this approach does not measure the actual skills of college graduates, nor does it estimate a specific contribution institutions make to those skills; instead, it uses wage as a proxy for skills (students who have higher wages are assumed to have more valuable skills). Considering graduate wages as a measure of college quality gained national attention in the recent discussions surrounding the gainful employment rules.<sup>40</sup> While focusing on students' ability to re-pay debt, these discussions have closely tied program performance to students' short-term labor market outcomes. Institutions can use wage data from the Bureau of Labor Statistics until 2015, at which point, they will have to use data from the Social Security Administration, which deal with individual students rather than averages for fields of study.

Although the gainful employment regulation has focused on programs at for-profit colleges and certificates and vocational programs at non-profit institutions, it has stimulated a broader discussion about “valuable” or “worthless” degrees. The National Governors Association’s March 2011 report *Degrees for What Jobs?* is perhaps the most visible of the recent endeavors to tie system and institutional performance to student labor market outcomes. The report narrowly defines the purpose of college in relation to labor market needs, and urges governors to demand that their higher education institutions develop courses and programs to prepare students for “high-paying, high-demand” jobs. Moreover, according to the report, public higher education institutions

---

<sup>40</sup> U.S. Department of Education. 2011. *Program Integrity: Gainful Employment-Debt Measures*. Available at: <http://www.federalregister.gov/articles/2011/06/13/2011-13905/program-integrity-gainful-employment-debt-measures>

should be required to collect and publicly report institutional impacts, assessed through indicators such as students' wages and employability.<sup>41</sup>

While paying attention to labor market outcomes of college graduates is potentially a useful enterprise, using graduates wage data as a measure to assess institutional quality presents many challenges. First, it is well known that school-to-work transitions in the U.S. are only weakly linked to educational institutions.<sup>42</sup> While colleges have placement offices and career services, institutional linkages in the U.S. are relatively weak, meaning that institutions have a limited ability to shape their students' labor market outcomes. Significant variation in outcomes is instead associated with student background, local labor market conditions and graduate geographic mobility. Research on institutional selectivity for example suggests that students' labor market outcomes have little if any relationship to institutions attended.<sup>43</sup> Moreover, none of the nationally representative large-scale datasets collected by federal agencies include measures of students' skills as well as wages. This makes it impossible to model empirically in comparative analyses the extent to which variation in skill development (e.g., test score growth) occurs across institutions relative to the extent to which graduate wage variation occurs across institutions. Such analyses would serve to either empirically demonstrate or call into question the technical feasibility of using graduate wage data as a metric of college performance at the institutional level. Some preliminary indicators can be obtained from the dataset collected for the *Academically Adrift* study,

---

<sup>41</sup> National Governors Association. 2011. *Degrees for What Jobs? Raising Expectations for Universities and Colleges in a Global Economy*. Washington DC: National Governors Association.

<sup>42</sup> Alan C. Kerckhoff. 1995. "Institutional Arrangements and Stratification Processes in Industrial Societies." *Annual Review of Sociology* 21:323-347.

<sup>43</sup> Theodore P. Gerber and Sin Yi Cheung. 2008. "Horizontal Stratification in Postsecondary Education: Forms, Explanations, and Implications." *Annual Review of Sociology* 34: 299-318.

but a more extensive consideration of this issue would require a different national data collection effort.

Second, focusing on early labor market outcomes can be largely misleading. Professional programs for example have higher initial wages, but lower earning trajectories.<sup>44</sup> This seems to in part reflect the advantage of the liberal arts and science majors in the process of promotion, particularly in the middle of the occupational hierarchy.<sup>45</sup> Moreover, many students choose their major with the intention to enroll in graduate/professional schools. This ‘option value’ of pursuing graduate education is greater for liberal arts and science fields, and students choose to major in those fields in part due to their expectations of continuing their education.<sup>46</sup> If graduate school enrollment does not immediately follow undergraduate education, or if graduate education is intertwined with employment (i.e., students take part-time or other less well-paying jobs while continuing their studies), early labor market outcomes of these graduates would reflect poorly on what in the long run may be an economically productive strategy.

Similarly, recent descriptions of certificate programs have highlighted the short-sightedness of trading immediate returns for a longer-term vision. While certificate programs are often able to place their students in jobs immediately after graduation, many of their graduates have little or limited ability to advance without additional education (and often more general education). Some institutions and states have considered

---

<sup>44</sup> Josipa Roksa and Tania Levey. 2010. “What Can You Do with That Degree? College Major and Occupational Status of College Graduates over Time.” *Social Forces* 89:389-416.

<sup>45</sup> Hiroshi Ishida, Seymour Spilerman, and Kuo-Hsien Su. 1997. "Educational Credentials and Promotion Chances in Japanese and American Organizations." *American Sociological Review* 62:866-882. Seymour Spilerman and Tormod Lunde. 1991. "Features of Educational Attainment and Job Promotion Prospects." *American Journal of Sociology* 97:689-720.

<sup>46</sup> Eric Eide and Geetha Waehrer. 1998. "The Role of the Opinion Value of College Attendance in College Major Choice." *Economics of Education Review* 17:73-82.

addressing this issue through “stackable certificates or degrees” – i.e., programs in which students acquire a lower-level certificate, which can be used to enter the labor market immediately; but these short-term certificates are then tied to degrees, which students can pursue subsequently to advance their economic position. Long-term outcomes of these students may be desirable, but it would be difficult to relate them to institutional performance of their certificate granting program. Judging school performance based on the outcomes of graduates from a decade earlier generates metrics that are not particularly useful for measuring current institutional performance and guiding ongoing reform efforts.

### Learning Outcomes

Instead of presuming that wages represent a certain level of skills, one could directly measure specific skills and competencies of college graduates. Indeed, measuring college performance by focusing on learning outcomes with value added test score growth has gained increasing attention. The Council for Aid to Education has promoted this strategy and facilitated its spread in the hundreds of institutions using the Collegiate Learning Assessment (CLA) to measure student learning.<sup>47</sup> Similarly, institutions participating in the Voluntary System of Accountability, sponsored by APLU and AASCU and including over 300 members, use a value-added strategy to assess student learning and report it publicly.<sup>48</sup>

The value-added strategy has many advantages over other approaches. First, the assessment strategy provides for more timely feedback and does not require post-

---

<sup>47</sup> See: <http://www.collegiatelearningassessment.org/>

<sup>48</sup> See: <http://www.voluntarysystem.org>

graduate tracking of individuals. Second, measurement is focused on an outcome that colleges can directly control and have implicitly assumed institutional responsibility for achieving. Third, the value-added methodology provides a relatively straightforward approach that is interpretable and aligned with normative values of equity and achievement: regardless of where individuals start in terms of performance levels, all students can learn and demonstrate growth. By focusing on gains in performance, the value-added approach is able particularly to self-adjust for individual-level differences: for example, if students are not motivated test takers or are subject to stereotype threat, their performance should be lower at each test administration, and schools should still be able to demonstrate gains in performance.

While a value-added focus on learning outcomes has much to commend it, what specific competencies are to be measured? Gaining reasonable consensus on learning outcomes is challenging enough, but developing assessments, particularly those that could be used across institutions, presents an even bigger set of challenges.

Although conceptualized in different ways, there is widespread agreement about three broad domains that potentially warrant assessment: generic higher order skills, subject specific knowledge, and affective growth/personal development. For logistical and pedagogical reasons, the primary focus in recent years has been on generic higher order skills (e.g., critical thinking and complex reasoning) as well as written communication. These competencies can be understood as the “21<sup>st</sup> Century skills” that are generally transferable across jobs, occupations and industries as well as are necessary for exercising responsible democratic citizenship. In addition, institutional mission statements and faculty surveys demonstrate organizational commitment to these

competencies.<sup>49</sup> Logistically, these competencies are empirically the easiest to measure in a value added longitudinal design as all students, regardless of curricular major, are expected to improve on these measures and performance can thus be assessed at college entry and then subsequently. There are several widely used assessment indicators that attempt to measure these generic higher-order competencies including the Collegiate Learning Assessment (CLA), the Collegiate Assessment of Academic Proficiency (CAAP) and the Proficiency Profile (formerly known as the Measurement of Academic Progress and Proficiency or MAPP).<sup>50</sup> While there are notable differences in the characteristics of these assessment tools, at the aggregate level they generate similar results.<sup>51</sup>

It is highly desirable that these assessments of generic higher order competencies are supplemented by measurement of subject specific knowledge. This is technically more difficult to accomplish in a longitudinal value-added framework as college students in the U.S. frequently drift in and out of majors throughout their collegiate enrollment. There are two ways to deal with this methodological challenge. First, short assessments could be given to all students taking introductory coursework in a particular subject area. For example, in the first weeks of an Introduction to Sociology course, all enrolled students could routinely be given a disciplinary focused skills and knowledge assessment regardless of planned major. These short assessments could be used as a basis to identify value added gains for students who subsequently went onto major in that subject area.

---

<sup>49</sup> Bok, *Our Underachieving Colleges*.

<sup>50</sup> Another test, the Critical Thinking Assessment Test (CAT), has been recently developed by researchers at Tennessee Tech, and funded by the NSF. Being more of a research than a commercial enterprise (although institutions can purchase it for their use), this test is less well-known and has not been incorporated into national efforts, such as the Voluntary System of Accountability.

<sup>51</sup> Stephen Klein, Ou Lydia Liu and James Sconing. 2009. *Test Validity Study*. Available at: [http://www.voluntarysystem.org/docs/reports/TVSReport\\_Final.pdf](http://www.voluntarysystem.org/docs/reports/TVSReport_Final.pdf)

This would require academic fields to articulate what competencies are being developed in particular majors – perhaps, a useful exercise given the growth of relatively weakly defined curricular programs on many campuses. An alternative, but less desirable methodological approach to measuring subject specific competencies would be to adjust student exit exams in a particular curricular area by controlling for measured student characteristics (e.g., social background, prior SAT/ACT scores, high school attended, etc.).

A third domain exists in terms of broader student development that is harder to measure effectively. This domain includes important components of affective growth and personal development. Researchers have with only limited success tracked individual growth in areas such as civic engagement, moral development, leadership skills, multi-cultural tolerance, creativity, etc. Nevertheless, this broad domain is normatively significant and should be acknowledged and recognized in assessment design. The Wabash National Study of Liberal Arts Education is one noteworthy recent endeavor that includes a number of indicators reflecting components of affective growth and personal development.<sup>52</sup> The Wabash Study, for example includes well-known measures of moral reasoning (the Defining Issues Test, DIT-2), leadership (the Socially Responsible Leadership Scale, SRLS-R2), and attitudes, cognitions, and behaviors regarding diversity (Miville-Guzman Universality-Diversity Scale – Short Form (M-GUDS-S)). By collecting longitudinal data on students from entry into college through their senior year, researchers using the Wabash data have been able to report value-added measures of these different indicators. One startling finding from the Wabash Study is that most of the affective and personal development indicators show smaller gains over

---

<sup>52</sup> See: <http://www.liberalarts.wabash.edu/study-instruments/>

time than do measures of generic higher-order skills, such as critical thinking.<sup>53</sup> Given the low average gains for critical thinking, and the general assumption by students and colleges that students are developing in other areas, even if not in the academic sphere, these results deserve careful attention in future research.

Considering the challenges of assessing students' affective growth and personal development, one possibility may be to shift methodologically from an individual-level value-added framework to an aggregate cross-sectional identification strategy for assessment of this particular domain. Consider, for example, the innovative accountability system adopted in New York City public schools, where one component of a larger assessment regime that primarily focuses on value-added test score gains, is based on student and teacher surveys measuring “learning climates”. In higher education, one could utilize similar student (and perhaps faculty) surveys – adapting existing instruments such as the National Survey of Student Engagement (NSSE) or questionnaires developed by the Higher Education Research Institute at UCLA – to identify the presence or absence of particular behaviors and institutional practices in this area. Although inadequate on its own, this strategy, when combined with other assessments, could potentially provide some insights into the extent to which schools foster climates that contribute to this broader aspect of student development. In order to allow comparability across schools, a set of assessment experts could be convened – as was the case in the New York City accountability design phase – that could provide input on a standardized set of questionnaire items that would be targeted to this domain.

---

<sup>53</sup> Charles Blaich and Kathleen Wise. 2011. *From Gathering to Using Assessment Results: Lessons from the Wabash National Study*. Champaign, IL: National Institute for Learning Outcomes Assessment.

*Conclusion*

As social scientists we should remind ourselves of our professional responsibilities as well as the limitations of contributing effectively to policy discourse on this topic. Our expertise can be usefully applied to identifying descriptively the normative dimensions of assessments and the technical feasibility of proposed metrics. What we are not in an appropriate position to do as social scientists is to advocate for what ought to be measured -- although as educators and citizens we have an obligation to develop value judgments and articulate principles about which goals we personally hold as desirable. “The distinctive characteristic of a problem of social *policy* is indeed the fact that it cannot be resolved merely on the basis of purely technical considerations which assume already settled ends,” Max Weber reminds us. Instead, “normative standards of value can and must be the objects of *dispute* in a discussion of a problem of social policy because the problem lies in the domain of general *cultural* values.”<sup>54</sup>

Social scientists could relatively easily and effectively design, evaluate and improve technical instruments that methodologically could assess college performance, if there was in fact a normative agreement among institutional stakeholders about what ought to be measured. “Goal setting is a political, and not a technical, problem,” David Labaree has suggested. “It is resolved through a process of making choices and not through a process of scientific investigation. The answer lies in values (what kind of schools we want) and interests (who supports which educational values) rather than apolitical logic.”<sup>55</sup> We believe that recent efforts to develop a consensus on such matters

---

<sup>54</sup> Weber, Max [1904] 1949. “‘Objectivity’ in Social Science and Social Policy,” in *The Methodology of the Social Sciences*, Pp. 49-112 in *Max Weber*, translated and edited by Edward Shils and Henry A. Finch. Glencoe, IL: The Free Press; p. 56.

<sup>55</sup> Labaree (1997); p. 40.

have been inadequate to the scope of these challenges. Well intentioned committees of academic professionals are not sufficient. Instead, political and institutional leaders have a significant role to play in identifying, articulating and rallying support for rationales that provide a compelling case for the specific functions of higher education today. Recent declines in state support for higher education suggest that existing institutional rationales have been inadequate to the task of maintaining resource flows. If higher education has come to be understood, not as a moral imperative, but rather simply as a system of allocating credentials for the labor market success of individual consumers, why should taxpayers and legislators feel compelled to invest scarce public resources in such an endeavor?

Focusing attention on the normative dimensions of college performance metrics recognizes that quantifiable evaluations and measurements in higher education are not new. In fact, we have been measuring features of college quality for decades, albeit with measures that arguably are not properly aligned with the normative commitments of many educators. It is the limitations of these existing measures and the problematic character of the organizational incentives that they have promoted that suggests the need for new metrics to guide institutional behavior. Rather than passively waiting for the dreaded imposition of externally imposed accountability measures, educators would do well to work proactively to make clear their normative commitments and support internal use of assessments aligned with their cherished values.