USING FEDERAL DATA TO MEASURE AND IMPROVE THE PERFORMANCE OF U.S. INSTITUTIONS OF HIGHER EDUCATION

September 2015

Contents

Introduction

Each year, tens of millions of Americans make choices about higher education. They decide whether to go to college, which school to attend, the level of education or training to pursue, and whether to enroll far across the country or close to home. Once enrolled in a higher education institution, students face another set of important choices, including whether to continue the program they're enrolled in, transfer to another, or leave their studies to immediately begin a career. To give a sense of scale, in the fall of 2013, roughly 20 million students were enrolled in 4,724 degree-granting institutions, up from 10.2 million students in 3,004 institutions in 1974. At the same time, there were another 472,000 students attending over 2,500 non-degree granting institutions that often involve shorter term certificate programs.[1] Uncounted in these numbers are the millions of individuals who considered college but opted not to attend, and the millions more who chose to leave their college studies.

Prospective students and their families must grapple with assessing which of the many institutions available will best prepare them to achieve their goals. The decisions to pursue a degree, which type, and at which institution all have a lasting impact on students and their futures. For example, obtaining a bachelor's degree can increase lifetime earnings by a million dollars relative to not pursuing postsecondary education (Carnevale, Rose, and Cheah, 2014), and evidence suggests that the 'quality' of the school attended can have large consequences for future outcomes. Similarly, differences in the price of attendance across institutions affect the financial returns, and may lead to differences in the amount that students have to borrow, which may also affect their career decisions and personal lives in meaningful ways. Despite the importance of this decision, there is a surprising lack of clear, easy to use, and accessible information available to guide the students making these choices.

College presidents and administrators, federal and state policymakers, and researchers are also increasingly interested in questions of institutional quality. College leaders need information on the performance of other institutions to better share and adopt best practices. Meanwhile, policymakers need better information to steward taxpayer dollars to schools that provide a good education to students and help students graduate with more opportunities. In the 2013-2014 academic year, state governments spent $76.2 billion in appropriations and grants to finance higher education, and the federal government provided $48.9 in financial aid and other educational benefits (including tax benefits) (College Board 2014a, 2014b). Reliable data is key to assessing institutional performance and determining how resources could be more efficiently allocated to benefit students.

Limited information about college outcomes partially reflects the complexity of measuring quality in higher education. Higher education institutions are diverse and provide a variety of benefits to both students and society as a whole. Since students have diverse goals—ranging from deepening their understanding of a field about which they are passionate to improving their

---

[1] *Digest of Education Statistics* (2014). https://nces.ed.gov/programs/digest/2014menu_tables.asp. These figures are for Title IV participating institutions only. See Tables 301.10 and 317.30.

core skills and competencies to finding a well-paying career or working in a particular occupation, and to other goals—and institutions cater to different interests and learning styles, an institution at the top of one prospective student's list may be at the bottom of another's. Similarly, a college that excels in one area, such as producing scientists responsible for important innovations, may lag in another, such as preparing students for a career in business.

Institutions that participate in the federal student aid programs must demonstrate that their students are able to repay their student loans under the cohort default rate and gainful employment provisions. These requirements intentionally focus on only one outcome of higher education – successful student loan repayment – and do not attempt to define or measure the broader purposes of higher education. In addition, the gainful employment rules focus on a category of vocational programs that are explicitly intended to lead to gainful employment in a recognized occupation and therefore can sensibly be evaluated on the basis of employment and debt outcomes alone. As a result, some of the issues raised in this paper may not be relevant for these existing accountability requirements.[2]

Efforts in the field are directed towards developing more comprehensive and summative measures of institutional quality. There is still value, however, in creating separate measures of various dimensions of college performance so that individuals and stakeholders can use actionable information to assess institutions according to their own priorities. Students care about which school will best prepare them to achieve their goals, whether that involves certain earnings after graduation or pursing a specific career path. Simple differences in student outcomes may be informative about differences in the institutions' contributions to students' success, but they may also be influenced by differences in the types of students who attend those schools. Regardless of the school they attend, students may have differing outcomes depending on their prior academic achievements, drive, interests, and other characteristics. Since schools attract and select different types of students, it may be hard to isolate the causal contribution of schools to students' outcomes – that is, how the individual student's outcome changed as a result of attending a particular college. Despite challenges in measurement, however, it is clear that students need the information available now to help them search for and select a college, and colleges need better information to benchmark and improve their performance.

To address the lack of information about college quality and costs, the Administration has created a new College Scorecard to provide reliable and unbiased information about college performance. Armed with this accessible and accurate information, students and their families will be able to make more informed decisions and better understand the consequences and

---

[2] The Department of Education has concluded that student characteristics are unimportant in the operation of the gainful employment rules because student characteristics exhibit a negligible relationship with the gainful employment metric (debt to earnings ratio) among applicable gainful employment programs; this is much less the case when comparing a broader set of outcomes across the full spectrum of higher education institutions. Moreover, in a case where some absolute level of an outcome is deemed the minimum acceptable performance level, policy makers should not consider using adjusted outcomes, since such an approach would likely permit worse outcomes for disadvantaged students. Adjusting such outcomes may also fail to disincentivize predatory recruitment practices that hurt students and divert taxpayer funds away from high-quality education programs.

tradeoffs of their choices. This report provides an overview of the new measures of college performance created as part of this effort and notes some important limitations in the data that should be kept in mind in their use. While most of these data do not necessarily reflect how a specific individual's outcome would change were he or she to attend a particular college, this report offers exploratory analyses of how federal data may be used to measure an institution's impact on a subset of performance measures. It also presents measures of college performance that might be of interest for accountability purposes and discusses several lessons learned from the data. While challenges remain, the College Scorecard data described in this report are a large step toward helping students and their families evaluate college choices; these never-before-released national data about post-college outcomes thus encourage colleges to strengthen supports that help students persist in and complete college, and to provide increased opportunities for disadvantaged students to get a college education. Moreover, further research by academics, discussions with stakeholders, and consultations with the public can take advantage of the increased transparency that these data afford to enable further progress in measuring college outcomes and the causal impact of colleges.

I.   Defining and Measuring Quality in Higher Education

The data described in this report reveal wide variation in the outcomes experienced by students attending different institutions of higher education. Figure 1-1 summarizes this variation for several key outcomes across two- and four-year institutions: The fraction of students who completed a degree at their institution, the fraction of borrowers successfully repaying their loan three years after entering repayment, and the median earnings of students who received federal aid who attended an institution 10 years after they began their studies. Each vertical bar in the Figure depicts the 10th-90th percentiles of the outcome on the x-axis: The bottom of the bar indicates the outcome (e.g., the completion rate, or median earnings among an institution's students) of the institution at the 10th percentile of institutional outcomes, and the top corresponds to the outcome of the institution at the 90th percentile. The darker shaded part of the bar ranges from the 25th to the 75th percentile—the interquartile range—and the line in the middle of that region indicates the value of the outcome for the median institution.

Figure 1-1: Distribution of Key Outcome Measures at 2 and 4 Year Schools

Note: Data are for the most recent cohorts. 150% completion rates from IPEDS, 3 year repayment rates, and 10 year earnings are used.
Source: Department of Education, College Scorecard Data
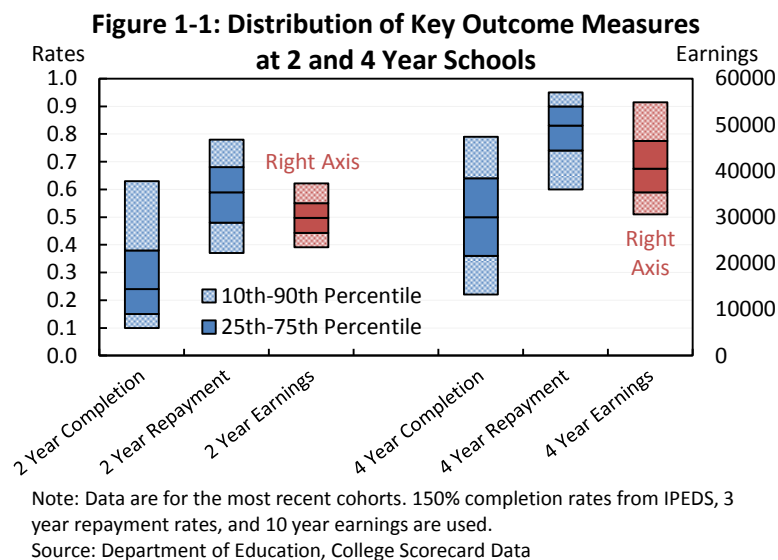
Figure 1-1 shows that outcomes for the typical student vary dramatically across institutions. For example, among four-year institutions, at the median institution, about half of students complete their degree. This is in contrast to the top 10 percent of four-year institutions, where more than four in five students complete their degree, and the bottom 10 percent, in which fewer than one in five students do so. Similarly, a typical student who attends a four-year institution at the top 10th percentile of the institutional earnings distribution has earnings of about $55,000 per year, while a typical student who attends a four-year institution at the bottom 10th percentile has annual earnings of about $30,000. If this difference remained constant, cumulated over a 30 year working career, a typical student at a top school would earn over $300,000 more than a student at a bottom 10th percentile school, in present value. In fact, this is likely an underestimate since earnings growth tends to be higher for workers with higher earnings.

The striking differences in outcomes across institutions suggest that many students may be making suboptimal choices in where to attend college, and more generally that many institutions are not generating the best outcomes for students.[3] This hypothesis is reinforced by the fact that Figure 1-1 shows large differences in student outcomes even among two-year colleges which tend to have open-admissions policies. An alternative explanation, however, is that the differences in outcomes may reflect differences in the types of students, for example in their academic preparation or interests, or in institutional mission. For instance, post-enrollment earnings may be low at a religious institution or school of music regardless of the quality of the institution. Conversely, at vocational colleges, earnings may more accurately reflect quality. Before evaluating institutions, it is important to understand the extent to which these differences in student characteristics or institutional goals drive variation in outcomes, as opposed to institutional quality. This section defines the term "quality" as it is used in this report, and briefly discusses some conceptual and empirical challenges in its measurement.

Defining Quality

A useful definition of *quality* for the purpose of this discussion is "the degree to which education services increase the likelihood of desired outcomes."[4] The essence of this definition is that quality is defined by the causal effect that attending an institution has on its students' outcomes, and not based on inputs such as the types of students it enrolls.

There are several broad challenges that make measuring quality in higher education difficult. As discussed below, one of those challenges relates to the difficulties in measuring the causal impact of an institution on student outcomes, given the pervasive sorting of certain types of students to certain institutions. It is worth emphasizing at the outset that while direct estimates of the causal impact of institutions on valued outcomes would be ideal, this does not imply that coarser information cannot or does not convey important quality information. Indeed, several studies in other contexts demonstrate that imperfect proxy measures of quality—even simple mean outcomes—can be useful guides for individual decision making (Hoxby and Turner, 2013; Kane and Staiger, 2008). Simple differences in average outcomes may be informative for some purposes, such as determining whether students may be able to repay loans based on the level of their debt relative to their earnings. In general, however, without credible estimates of institutions' causal effects on outcomes, it is difficult to understand how differences in average

---

[3] Of course, not all institutions are accessible to all students as their choices are limited in a variety of ways. In addition to obvious considerations of cost and location, many schools are accessible only through a competitive admissions process that will not allow certain students to choose institutions attractive to them that may offer better results. Moreover, many institutions, especially those with high outcomes, do not expand their enrollment to accommodate all qualified applicants (in part due to incentives to maintain their selectivity). Thus, while better information can help students make more optimal decisions among the institutions open to them, admissions constraints mean that they may not necessarily be able to move to some of the institutions with the highest outcomes described in this report. Importantly, however, the data reveal wide variation in outcomes even among less- and non-selective schools within the same region.

[4] This paraphrases a definition from the Institute of Medicine (1990), which also adds the process oriented condition "… and are consistent with current professional knowledge." While potentially important, process factors such as pedagogic practices, etc. are not considered in this report.

outcomes indicate quality differences, motivating the exploratory analyses discussed in section IV of this report.

II. Information and Decision Making in Postsecondary Education

A natural framework to consider the types of information most useful to students is to view students as making choices between colleges (or attending college versus not enrolling at all) that attempt to maximize their expected net benefits. To do so, prospective students need information on the potential outcomes and financial costs associated with alternative choices, conditional on their interests and talents. Several features of this framework are important in considering the types of information students need:

1. The set of outcomes and college attributes relevant for college choice is likely to be extremely large, since students have heterogeneous goals related to higher education. As a result, any source of college information will be an incomplete guide for some students who will need to use outside sources to find information on their particular interests. As discussed below, however, there is a core set of information that most students indicate is very important to their decisions.

2. Prospective students assign different relative importance to expected outcomes and college attributes, so presenting combined summary measures of several individual outcomes may obscure important information.

3. Cost of attendance may differ dramatically based on family income, academic background (or other characteristics such as athletic or artistic prowess), and financial aid availability. Each student cares about what he or she will likely need to pay, so providing customized information is important.

4. Finally, student outcomes at each institution may vary with student characteristics such as family income or academic interests. Customized information about differences in outcomes for students with their own characteristics can be helpful to students. This means examining each institution's contribution by asking, what are the outcomes for students with similar characteristics across different institutions?

In addition to providing information to prospective students and their families, there is growing demand from federal, state, and local governments for information about the performance of higher education institutions and to know whether investments in these institutions are paying off. While these actors likely value the same information as potential students because they value the net benefits of education to students, they are also likely to put value on a broader set of benefits and costs generated by colleges. These are discussed in greater detail below.

Information to Guide Student and Family Choices

Potential Students Have Varied Goals and Information Needs
Students pursue higher education with a variety of goals in mind. A UCLA review of incoming freshman at four-year colleges found that the most reported reason to attend college (at about

85 percent of students) was the ability to get a better job, and about 75 percent of students stated that getting training for a specific career was also very important. The second most common answer, indicated by over 80 percent of students, was learning more about areas of interest; and over 70 percent stated that gaining a general education and appreciation of ideas in college was very important too (Eagan et al., 2014). Similar data from a poll by New America found that while roughly 90 percent of students go to college to improve employment opportunities, make more money, and get a good job, over 80 percent also aim to "learn more about a favorite topic or area of interest" and "become a better person" (Fishman, 2015). Over half of the respondents to the New America survey rated the following reasons as important or very important reasons to attend a specific school: The majors or programs offered, availability of financial aid, cost, location, number of graduates employed full-time in field within six months, graduation rates, starting salaries for graduates, average amount borrowed, and loan default rates. [5]

Public college attendees weight affordability and location as key reasons for enrolling in a specific institution, while private, nonprofit four-year students weight reputation and location most highly (Radford, Tasoff, and Weko, 2009). Students from different demographic groups also may have different priorities. Older students are often driven by career change or advancement goals while younger students' career goals relate to entry into a field of choice (National Postsecondary Education Cooperative, 2007). Distance and cost also matter more for older students and those from disadvantaged backgrounds (Eagan et al., 2014; National Postsecondary Education Cooperative, 2007).

Given the heterogeneity in the goals and priorities of students, providing a comprehensive rating of institutions that incorporates information about their goals, weighted in a way that matches their priorities, is challenging. Instead, providing clear and transparent information about the costs and attributes of college is both important and feasible. It is important that students and families clearly understand the costs they can expect to incur through their studies, the amount of loan payments they will have to make if they finance the cost of attendance through borrowing, and a sense for whether they will be successful in attaining a degree and in the labor market upon leaving the institution. These factors may not be the most important criteria to students and their families, but it is important that the information exist so it can inform choices about whether and where to attend college. Armed with detailed, relevant information on financial costs and benefits, students can more fairly evaluate the tradeoffs of attending a certain institution and understand the financial implications of their decisions.

How Might More and Better Information Help?
A large and growing body of literature shows that individuals considering whether and where to attend college currently have too little information about costs and outcomes. Providing more and better information to consumers can measurably improve prospective student outcomes.

---

[5] The New America polling sample consists of representative 1,011 U.S. residents ages 16-40 planning to attend college in the next 12 months or who are in their first semester of a two- or four-year college.

Information about Affordability

College costs are hard to accurately assess, especially for low-income parents who are less likely to have attended college themselves. Avery and Kane (2004) find that low-income and first-generation college students tended to overestimate the cost of college and failed to take steps in the application process that would reduce this cost, such as filling out financial aid forms or applying for application fee waivers. Other studies corroborate the difficulty that low-income parents and students have in accurately estimating costs (Grodsky and Jones, 2007; Horn, Chen, and Chapman, 2003). Much of this behavior appears to be driven by confusion over federal, state, and institutional financial aid that can reduce the tuition paid by a student below the advertised rate. A survey by the College Board (2010), for example, found that 59 percent of students looked only at stated levels of tuition (sticker price) in evaluating the cost of attending a school, without taking into account financial aid like Pell Grants and school-provided aid. This may affect if and where students ultimately decide to enroll in school, given that students who overestimate costs are less likely to matriculate in any degree program, believing the price is higher than it actually is (Hastings et al., 2015).

Multiple studies have found that increasing the amount of information students receive about financial aid has significant effects on application and enrollment rates. For instance, Bettinger et al. (2009) found that students who received assistance filling out their FAFSA and personalized aid estimates were significantly more likely to file the FAFSA (15.7 percentage points more than students in the control group, or 40 percent) and enroll in college (29 percent). For students from low-income families, the college enrollment impact was even greater (40 percent).[6]

Similarly, Hoxby and Turner (2013) evaluated an intervention, the Expanding College Opportunities (ECO) project, intended to improve the information that high-achieving, low-income students have when applying to college. The study tested several interventions to assess why high-achieving, low-income students do not apply to more selective schools, including providing application guidance to find schools that were a good academic fit, offering information about the differences between net and list prices, and offering fee-waivers for nearly 200 selective colleges. The study found that the combined effect of all of the interventions was to cause targeted students to apply and gain admission to more selective institutions and ultimately to enroll in institutions with higher graduation rates and higher levels of student spending—both measures of school quality made available to students. While the study found that information about net price alone may not—unsurprisingly—lead students to ultimately enroll in higher quality (or better academic fit) institutions, the pattern of results for each intervention suggests that both the "net cost" and "application guidance" portions increase the number of applications students send.

Information about Earnings, and Other Student Outcomes

There is growing evidence suggesting the information that students have about the extent to which college choice might lead to different outcomes, particularly in the labor market, is poor. Economic literature suggests that uncertainty in the returns to college can deter many students

---

[6] The absolute increase was nearly 7 times smaller for low-income families, but relative to a much smaller baseline enrollment rate the increase was proportionally larger.

from attending college, even those who would have benefited financially (Heckman, Lochner, and Todd, 2006). Students from lower-income families and those who are first-generation college students have even less access to information and are more concerned with cost (National Postsecondary Education Cooperative, 2007).

Key outcome measures that students and their families might consider include graduation and transfer rates, measures of labor market success like earnings and employment, and a sense for how much they will need to pay for their education, including the amount of loan payments they will need to make after completing their studies and the likelihood they will be able to afford these payments. Armed with detailed, relevant information on financial costs and benefits, students can more fairly evaluate the tradeoffs of attending a certain institution and understand the financial implications of their decisions.

Research shows considerable variation in the earnings that high-school and undergraduate students associate with various levels of education and experience (Betts, 1996; Dominitz and Manski, 1996). One study found that college students are misinformed about population earnings, but that they revised their earnings expectations and retained that information after being provided with accurate details on education and earnings potential (Wiswall and Zafar, 2013). Another study found that providing students with median earnings and earnings variability led to different expectations and major choices for students (Ruder and Van Noy, 2014). Both uncertainty about and systematic underestimation of the returns to schooling may lead to students' underinvestment in education.

Jacobson and LaLonde (2013) present data suggesting that insufficient information about student outcomes is a pervasive problem in community colleges as well. In particular, they document that many potential trainees underestimate the returns to career-oriented or vocational programs, some of which can lead to earnings nearly 50 percent higher than degrees in lower-return fields. Jacobson and LaLonde suggest that individuals who can benefit from training most frequently have poor access to reliable information from friends and relatives on returns, and that training institutions devote too small a fraction of their resources to helping trainees make optimal choices about which program is right for them.

Information to Guide Accountability Efforts

A related but distinct use for college performance measures is to enable "consequential accountability" policies (referred to in this paper simply as accountability). These systems tie funding or increased oversight to performance outcomes.[7] Meaningful accountability efforts are focused on encouraging colleges and universities to focus on outcomes for students from a diverse range of backgrounds rather than just inputs such as enrollment, to drive improved

---

[7] Hanushek and Raymond (2005) define the terms "consequential" and "report-card" accountability in an overview of state K-12 policies. "Report-card" accountability refers to systems that rely on providing information to stakeholders, whereas "consequential" accountability schemes attach consequences in the form of funding changes or other oversight to performance outcomes. In this report, we refer to this notion of "report-card" accountability as "public accountability."

performance. While there is a large degree of overlap in the information that is relevant to consumers and the measures that are relevant for accountability efforts, there are at least three important differences: 1) accountability systems may value a different or a broader set of outcomes than those that may be directly relevant to consumers; 2) government ought to be concerned with the overall resource cost beyond the prices paid by students; and 3) some forms of consequential accountability schemes require an explicit mapping from various dimensions of performance to consequences, and therefore an explicit scheme for valuing the relative importance of each dimension of performance.

First, accountability systems that aim to increase the overall social value created by public investments in postsecondary institutions may value a broader set of outcomes from institutions of higher education than those valued by individuals. Research shows that increases in postsecondary education levels have spillover benefits to society that may not be fully captured or valued by the individual receiving an education. These benefits can accrue in part to other members of society potentially through reduced victimization and lower health care and law enforcement costs (Baum, Ma, and Payea, 2013; Hill, Hoffman, and Rex, 2005; OECD, 2013). Similarly, government might also value social contributions of individuals—such as voting, volunteering, teaching, inventions, or public service—not fully captured by the individual's wages.[8] These types of outcomes can be extremely hard to measure, however, and there is little consensus or evidence regarding which such outcomes are most relevant.

Second, accountability systems may incorporate information on enrollment of low-income or other underrepresented groups, in order to create incentives for institutions to expand or preserve access to those students. As shown in Figure 2-1, a child who attains a college degree is much more likely to move to a higher income quintile in adulthood than a child who does not earn a degree (Haskins, 2008). So while there is still an unfortunately large gap in college access between individuals born to high- and low-income families, with an even larger gap for completion, low-income students who attend college do see overall increases in future income as opposed to low-income students who do not.

---

[8] See, for example, Lockwood, Nathanson, and Weyl (2012) and Baumol (1990) for discussions of how net externalities generated by some professions might affect socially optimal incentive structures.

**Figure 2-1: Social Mobility - Adult Children's Income Quintile
for the Poorest Quintile of Parents, by College Degree**
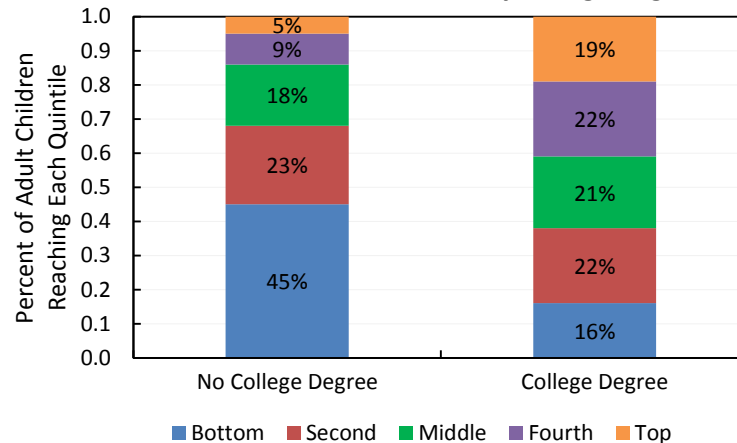
Source: Brookings Tabulation of the Panel Study of Income Dynamics data

In addition to valuing a broader set of outcomes, government may want to create incentives for institutions to reduce costs. State and federal government grant aid defray the list price to students, allowing most students to pay less to attend college. But these resources are limited, and governments need to be concerned with whether institutions are using these funds efficiently to produce outcomes for their students. While measurement of total school expenditures on the set of activities that may influence students' outcomes is difficult, researchers proxy this by using either instructional spending per full-time equivalent (FTE) student or education and related expenditures per FTE student (Desrochers and Wellman, 2011). The latter measure is considered the "most robust measure of spending on student learning" and includes spending on instruction, student services, and an allocated portion of general support and maintenance costs associated with these functions (Desrochers and Wellman, 2011, p20).

A final distinction is that using college performance information for high stakes decisions generally requires a way of combining various measures of performance into a single index of school performance. Higher values of the index could be tied to rewards, such as higher funding, and lower values of the index could be tied to sanctions, such as ineligibility or reduced eligibility for student or institutional aid. Creating such an index involves an explicit numerical weighting of various outcomes, requiring normative judgments about the relative importance of different dimensions of institution performance.

An Example: State Performance-Based Funding
An increasing number of states are adopting accountability policies that tie institutions' funding to the outcomes they produce, sometimes called "pay-for-performance" systems. These performance based funding (PBF) policies provide incentives for colleges to produce student outcomes in line with public goals. As of 2014, 32 states had such systems in place, with five more in the process of creating a system. Five of the 30 states use pay-for-performance only for four-year schools, while five states use it for only two-year schools (National Conference of State Legislatures, 2015). Most states set aside between five and 25 percent of higher education dollars

for performance funding, but this can vary greatly, from less than one percent in Illinois to virtually 100 percent in Tennessee (National Conference of State Legislatures, 2015; Jones, 2013). In general, states are moving towards increasing the share of funding based on performance metrics (Jones, 2013; Snyder, 2015).

Performance-based funding models are a good illustration of the broader demands for performance information that characterizes accountability systems. Table 2-1 below summarizes the institutional performance measures considered in various states' PBF schemes. Student progression toward completion and degree attainment measures are featured in most states' systems. Currently, 28 states consider the number of degrees awarded by a university, 16 use some form of course completion, 12 include retention rates, and 12 incorporate graduation rates. Another 11 states put weight on post-graduation outcomes such as job placement rates, licensure test passing rates, or earnings.

Many states clearly attempt to motivate institutions to pay extra attention to outcomes for certain student subgroups, reflecting the value placed on institutions' contribution to social mobility. For example, in more than half of states with PBF, special weight is given to either the completion outcomes for or enrollment of Pell-eligible or other-defined low -income, minority, first-generation, at-risk, non-traditional, transfer, or "older" students. These weights can serve an important purpose when the metrics have potential to create perverse incentives. For example, while states would like to improve institutions' graduation rates, they would not like colleges to meet this target by reducing the number of low-income students they admit. To that end, they give extra weight to the graduation rates of students who tend to experience worse outcomes. States also place value on aspects of college performance that are not directly related to student outcomes. Fourteen states reward research and development-related activities like the number of patents generated or external funding obtained, and some states also consider faculty outcomes like tenure and advancement, partnerships with K-12 schools, and faculty diversity.

Table 2-1: Metrics for Accountability by State		
Metric	**Number of States**	**States**
Any Performance Based Metric	**34***	AZ, AR, CO*, FL, GA*, HI, IL, IN, IA*, KS, LA, ME, MA, MI, MN, MS, MO, MT, NM, NV, NC, ND, OH, OK, OR, PA, SD*, TN, TX, UT, VA, WA, WI, WY*
Degree Numbers	**28**	AZ, AR, CO, GA, HI, IL, IA, KS, LA, ME, MA, MI, MN, MS, MO, MT, NV, NM, OH, OK, OR, PA, SD, TN, TX, VA, WA, WI
Overall degrees or awards	26	AR, CO, GA, HI, IL, IA, KS, LA, ME, MA, MI, MS, MO, MT, NV, NM, OH, OK, OR, PA, SD, TN, TX, VA, WA, WI
Degrees or awards per FTE	7	IL, ME, MA, MS, NV, TN, VA
STEM degrees	12	AZ, AR, HI, IL, KS, ME, MN, MS, NV, PA, SD, TX
Degrees for Pell recipients	4	IL, ME, MA, MS
Degrees for low income students	3	IL, NM, TN
Degrees for one or more of minority, first generation, at-risk, adult, non-traditional, or transfer students	7	GA, IL, IN, ME, MS, OR, TN
Growth in degrees granted	3	AZ, IN, MN
Graduation Rate	**12**	AR, CO, FL, IL, IN, KS, MI, MN, MO, NC, TN, UT
Overall graduation rate	3	IL, MN, UT
Graduation rate for Pell recipients	2	AR, CO
Graduation rate for low income students	1	MN
Graduation rate for one or more of minority, first generation, at-risk, non-traditional, or transfer students	2	AR, CO
On-time graduation rate and/or growth in on-time graduation rate	1	IN
Three year graduation rate (for two-year institutions)	2	KS, MO
Six year graduation rate (for four-year institutions)	6	FL, KS, MI, MO, NC, TN
Enrollment	**17**	AR, CO, FL, HI, IL, IA, MA, MI, NV, NC, OH, PA, TN, TX, UT, VA, WI
Number of minority students enrolled	4	IA, OH, PA, VA
Number of Pell recipients enrolled	6	AR, FL, HI, MI, OH, PA
Number of low income students enrolled	1	IA
Number and/or percent of students who transfer from community colleges	13	CO, HI, IL, IA, MA, MI, NV, NC, OH, TN, TX, UT, VA
Number of dual enrolled (high school) students	4	MI, TN, VA, WI
Retention & Course Completion	**27**	AZ, AR, CO, FL, GA, IL, IN, IA, KS, LA, MA, MS, MO, MT, NV, NM, NC, ND, OH, OK, TN, TX, UT, VA, WA, WI, WY
Overall retention and/or progression rate	12	AR, CO, FL, GA, KS, MO, MT, NC, OK, UT, VA, WA
Retention rate growth	2	CO, LA
Student credit hours	12	AZ, IA, MA, MS, MO, NM, ND, OH, OK, TN, TX, WA
Course completion and/or course completion rate	4	AR, NM, OH, WY
Completion of certain math and English course levels	9	MA, MS, MO, NV, NC, OH, TX, UT, WA
Remedial and/or adult education advancement course credits	5	AR, IL, IN, TN, WI
Workforce training course credits or course credits with industry-validated curricula	2	AR, WI
Post-graduation Results	**11**	AR, FL, IA, KS, LA, MO, MN, NC, TN, UT, WI
Job placement or continuing eduction rates and/or rate increase	8	AR, FL, IA, KS, LA, MN, TN, UT
Job placement related to students' programs of study	1	WI
Post-graduation wages	2	FL, KS
Licensing exams pass rate	3	LA, MO, NC
Expenditure	**9**	FL, IL, MS, MI, MO, MN, PA, UT, VA
Tuition and/or cost per credit hour and cost per completion	2	IL, VA
Tuition increase	1	MI
Administrative spending and/or spending changes	3	MI, MN, PA
Cost per student or degree to the university	4	FL, MS, MO, UT
Institutional support as a percent of core expenditures	1	MI
Private support dollars raised	1	PA
Research and Development	**14**	AZ, AR, IL, IA, ME, MI, MN, MS, NV, NM, SD, TN, UT, VA
Research and development and/or public service external funding received or applied for	6	AZ, AR, ME, NM, TN, UT
Research and development and/or public service expenditures	6	IL, IA, MI, MS, NV, VA
Research and development expenditure growth	1	SD
Number of patents or invention disclosures	4	AR, MN, MS, VA
New company start-ups	1	AR
Selectivity	**2**	FL, KS
Rank on a regional or national ranking	1	KS
Performance on quality measures compared to peers	1	KS
Freshman in top 10% of graduating high school class	1	FL
Other	**3**	LA, PA, VA
Institutional efficiency and accountability	1	LA
Faculty diversity	1	PA
Faculty career advancement	1	PA
Campus safety and security	1	VA
K-12 partnerships	1	VA
Note: * denotes that CO, GA, IA, SD are currently transitioning to performance funding programs		
Source: National Conference of State Legislators, 2015 (http://www.ncsl.org/research/education/performance-funding.aspx)		

Tennessee's outcomes-based funding formula for four-year institutions offers an illustrative example of the specific choices a state must make. Tennessee state evaluates universities on several outcomes, almost all of which are student focused. While complicated, a glimpse of the details helps to illuminate the difficult set of choices that must be made in any public accountability system. The formula used to map outcomes on these measures to state funding for four-year universities is

$$
\begin{aligned}
F = \Big[& w_{H24}(H24 + 0.4H24_A + 0.4H24_P) + w_{H48}(H48 + 0.4H48_A + 0.4H48_P) \\
& + w_{H72}(H72 + 0.4H72_A + 0.4H72_P) + w_D(D + 0.4D_A + 0.4D_P) + \frac{w_M M}{0.3} \\
& + \frac{w_L L}{0.05} + \frac{w_R R}{20{,}000} + w_T T + \frac{w_Q Q}{0.02} + \frac{w_G G}{0.04} \Big] \times S
\end{aligned}
$$

where F is a funding amount determined by the outcomes measured; H24, H48, and H72 are measures of the number of students accumulating various credit hour benchmarks; D is the number of bachelor's and associate's degrees awarded; M is the number of master's degrees awarded; L is the number of doctoral or law degrees awarded; R is the total amount of external research and grant funding; T is the number of student transfers out; Q and G are measures of degree production and a six-year graduation rate, respectively; and S is a measure of average faculty salary. The subscripts 'A' and 'P' denote outcomes measured separately for students over the age of 24 and for low-income (Pell-eligible) students. The weights 'w' are predetermined by the state and vary within a range chosen based on the institution's mission—for research universities the state assigns more weight on graduate degree outcomes and research, whereas for teaching schools relatively more emphasis is placed on undergraduate student progression and degree attainment. An analogous funding formula for two-year schools reflects the different mission of those schools, for example including measures of workforce training, remedial and developmental success, and dual high school enrollment, but ignoring graduate degrees and research expenditures (Tennessee Higher Education Commission, 2015).[9]

The formula helps to illustrate the three distinct aspects of an accountability system described above. First, the Tennessee model places extra value on the credit accumulation and degree attainment outcomes of both adult students (those over the age of 24) and Pell-eligible students. In particular, these outcomes are valued 40 percent more than the same outcomes for traditional age, non-Pell eligible students. The intent of this adjustment is to give institutions incentive to enroll and promote the success of "older" and lower-income students. An interesting facet of using this "bonus point" approach, as opposed to relying on regression-adjusted outcome measures (discussed in Section IV), is that the incentives of the institution to enroll disadvantaged students can be varied with the choice of the "bonus factor" (i.e., the 0.4 additional weight in the formula). If, for example, Pell students' degree attainment rates were exactly 71 percent (1/1.4) as high as non-Pell eligible students, then the formula above could eliminate an institution's incentive to not enroll Pell students (abstracting from other impacts on other outcomes, or cost

[9] Further details about the funding formula are available at http://www.tn.gov/thec/topic/funding-formula-resources.

considerations). If Pell degree attainment rates were less than that, institutions would still have a disincentive to enroll them (that could be avoided by increasing the bonus factor), whereas if Pell attainment rates were more than 71 percent of non-Pell students then the 40 percent bonus would incentivize schools to enroll more of them. At a broader level, the formula illustrates how systems can attempt to both reward strong student outcomes and promote access for all students.

The formula also shows that some measures not directly of relevance to students can be included, though Tennessee's funding model is heavily oriented around student outcomes. In particular, measures of external awards are incorporated in the model. Finally, the formula illustrates well how various dimensions of institutional performance can be combined into a single index of performance, in this case the level of funding. Note in particular how different measures of performance are weighted in the formula. For example, the formula suggests that in teaching schools (where w_R and w_T are both 10 percent) if every student successfully transferred to another institution (T) the state would deem that to be equivalent to gaining $20,000 in research funding and would reward it accordingly with state support.

III. Using Federal Data to Measure Access, Affordability, and Student Outcomes at Postsecondary Education Institutions: Available data and Key Considerations and Limitations

Simple descriptive data, along with statistics about the types of students enrolled at different institutions, provide easy-to-understand information that students can use to assess school quality. While more sophisticated techniques to assess institutions' contributions to student success are discussed in the next section, this section provides an overview of the array of data maintained by the federal government that can be used to measure the performance of postsecondary education institutions in terms of the access they provide, the level of affordability, and the outcomes of students. Some of these sources are already widely used by the higher education community. However, this report also documents several newly created measures of college performance based on large administrative datasets maintained by the Department of Education and the Department of the Treasury. This section highlights the more important measures and their data sources and discusses some important limitations and considerations of which users of this data should be aware. More detailed notes on the construction of these measures are provided in the Appendix.

Data Sources and Performance Metrics

Integrated Postsecondary Education Data System (IPEDS)

Collected annually through surveys administered by the Department of Education's National Center for Education Statistics (NCES), IPEDS is the primary source of data on postsecondary education institutions in the United States. The data include important performance information such as graduation rates for student subgroups; retention rates; tuition, cost of attendance, and net prices; and enrollment of low-income students (i.e., the number of Pell Grant recipients). The data also include a host of institutional information including location, type of control, level of degrees offered, enrollment size, institutional resources (including faculty-per-student ratios and faculty salaries), institutional selectivity, and characteristics of enrolled students, such as the 25th and 75th percentiles of admissions test scores for students who submitted their results.

Under the *Higher Education Act*, all institutions that participate in Title IV federal student aid programs must complete the IPEDS questionnaires. As such, IPEDS provides a broad starting point to define a universe of institutions on which to focus attention. In the most recent collection in 2013, there were 7,253 institutions in the 50 states and Washington, D.C. that were active participants in the Title IV program. We further limit the scope of analysis to institutions that served at least some undergraduate degree-seeking students and awarded some undergraduate degrees or certificates in the 2012-13 school year to form the initial universe of institutions for the data presented in this report.

As discussed below in more detail, IPEDS data have several important limitations for measuring institutional performance. Perhaps the most significant is that many outcomes are recorded for a limited subset of students. Most importantly, graduation rates are only reported for cohorts of

first-time, full-time students, so graduation rate information is not available for students who may have previous higher education experience, including transfer students, or for part-time students.[10] Another limitation is that outcomes are not recorded for students who transfer from the institution. Thus, information on graduation rate outcomes—important both for institutions that prioritize preparing students for a four-year degree as part of their mission and for students considering which institutions will help them succeed on such a path—is limited.

While this report uses many IPEDS elements, some outcomes of interest related to institutional performance are:

1. Cohort graduation rates for first-time full-time students. This is the official measure of graduation rates mandated by the Higher Education Act, measuring the fraction of first-time full-time students who complete their program of study within 100, 150, or 200 percent of the 'normal' completion time—e.g., the 150 percent completion rate measures the fraction of the cohort that graduates within six years for students pursuing a four-year degree or three years for students pursuing a two-year degree. While this measure is standard, it has been criticized for covering a small fraction of the student body at some colleges. For example, at community colleges, the share of the student body that is first-time full-time is often a small minority of the total students enrolled.

2. Average net price and cost of attendance information for first-time full-time students who receive federal financial aid. Since 2009, IPEDS has asked institutions to report information about tuition, fees, and other living costs that together constitute an estimate of the total cost of attending the institution. Institutions also report the average net price for any student receiving grant or scholarship aid from federal, state or local governments, or from the institution. Average net price can then be computed for both this group overall or for Title IV students only. While this is the best available metric for measuring average institutional costs, it is influenced both by the family income backgrounds of those who enroll, as well as by the availability of aid to students. More accurate affordability information is available by measuring the average net price for students across five income categories for Title IV recipients. A limitation of all of the measures is that for public institutions, the measure only captures the net price for students paying in-state (or in-district) tuition.

3. The fraction of enrolled students who receive a Pell grant. This is one of the most commonly used measures of the degree to which institutions provide access to low-income students. For example, about 75 percent of financially dependent Pell recipients are from families with family income below $40,700; the same share of independent students had family income lower than $23,300.[11] A limitation of this measure in reflecting the extent to which the institution serves low-income students is that many

---

[10] There is information on total numbers of degrees awarded to students of various types, but this cannot be used to compute outcomes for enrollment cohorts.
[11] Based on CEA calculations using the 2012 NPSAS, reported in real 2014 dollars.

low-income students never apply for aid or may not be eligible (e.g., undocumented and international students), and so are not captured in the measure. Additionally, the fraction receiving Pell makes arbitrary distinctions between students from families with very similar family resources, as students with family incomes just above the Pell eligibility thresholds are not counted in the measure. It may also create the undesirable incentive for institutions to substitute away from the "near poor" students who are barely ineligible for Pell.

National Student Loan Data System (NSLDS)

The National Student Loan Data System (NSLDS) is the Department of Education's central database for monitoring federal student aid—primarily federal student loans and Pell grants. The NSLDS contains data exclusively on federal borrowers and grant recipients dating back to the 1960s. It is used primarily for operational purposes, such as tracking federal grant and loan disbursements, the enrollment status of aid recipients to determine repayment status, payments on federal loans, borrower status (e.g., deferment, forbearance, or default), and remaining loan balances.

Since the late 1980s, NSLDS data have been used to produce the institutional cohort default rate (CDR), the only accountability measure currently used to determine eligibility for Title IV participation for all schools. As of July 2015 it is also used to estimate total accumulated loan balances, which are used to create the debt to earnings measures that determine Title IV participation for programs under the gainful employment rule.[12] The data system also contains a host of other information that can help inform prospective students and policymakers about colleges' performance. Of course, additional information exists about loan balances and repayment behavior over the lives of all borrowers. This includes the debt that students accumulate while in school, including originated amounts by type of loan and the balance owed on both principal and interest at various points in time. These data can be used to produce a variety of new institutional performance metrics that are described and presented below:

1. The median cumulative loan debt originated at the institution for all student borrowers of federal loans who leave the institution (i.e., either graduate or withdraw) in a fiscal year, measured at the point of separation. This gives a measure of how much a typical borrower borrows to attend the institution and thus can provide students with a sense for how much to save in preparation to attend and how much they will need to earn to cover their debt service payments after they leave college. Note that this measure does not reflect borrowing for the typical student (including non-borrowers), since the fraction of students who borrow varies greatly across institutions. Thus, information from IPEDS on the fraction of undergraduate students who borrow thus provides important context for inferring the level of borrowing for the typical student at the institution overall. Importantly, this measure is also available for the set of borrowers who complete their

---

[12] Under the gainful employment rule, the estimated annual loan payment of a typical graduate may not exceed 20 percent of her discretionary income or 8 percent of her total earnings (over several years) in order to be eligible for federal student aid programs (Department of Education, 2015).

degree. This allows a more accurate characterization of the total borrowing a student should expect if they complete their studies at an institution.

Figure 3-1: Distribution of Median Total Debt for Graduates Across Institutions
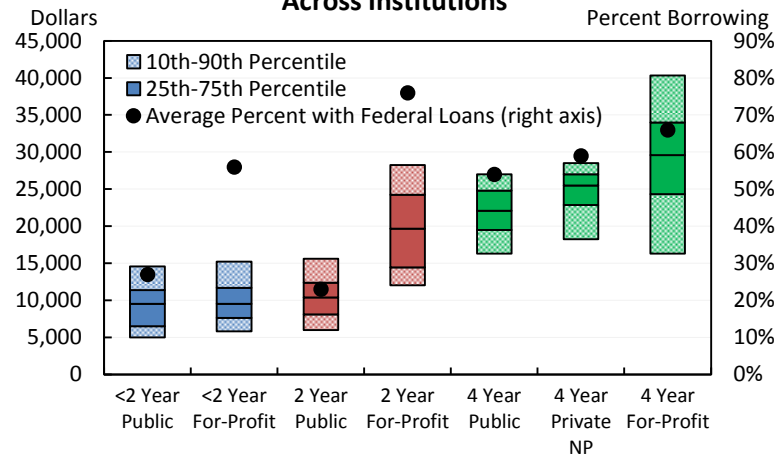
Note: Debt data are for the 2013 and 2014 pooled exit cohorts.
Source: Department of Education, College Scorecard Data

2. <u>Institutional cohort repayment rate</u>, or the fraction of student borrowers who are making at least some progress paying down their loans (i.e., their balance is declining) and are not in default. This measure is conceptually similar to a cohort default rate but is intended to be less susceptible to artificial manipulation, which may occur if an institution pushes students into deferment or forbearance until the measurement window expires. The repayment rate is measured at one, three, five, and seven years after entering repayment.

3. <u>Cohort completion and transfer rates</u> for all federally-aided students who begin their education at an institution in a particular award year. Since aid disbursements are associated with institutions, NSLDS data can be used to identify the set of students who are first aided in a particular award year at each institution. While some students may first receive federal aid after their first year of enrollment, student-reported data on grade level from the student's Free Application for Federal Student Aid (FAFSA) can be used to place students in a cohort based on the year in which they first enrolled in a college. Since institutions report when students graduate or leave school, cohort completion rates can be estimated for each institution. Moreover, transfer to and completion at other institutions can be measured for the same cohorts by tracking whether students enroll and/or graduate at other institutions. As discussed below in the "Important properties and limitations of federal data" section, since NSLDS has not traditionally been used to measure these outcomes, there are several important caveats to these measures. Currently these data are intended to be used by researchers and institutions to gauge and benchmark their performance relative to other institutions and to help generate better understanding of the validity of these metrics. Until their validity are better understood, the data should not be used for consumer information applications, as they have the potential to mislead.

23

4. <u>Student demographic information</u> taken from FAFSA forms. While IPEDS provides information about prior achievement, racial and ethnic composition, and other aspects of enrolled students, data from the FAFSA complement that information in several ways. First, since the data cover all federal aid recipients, the FAFSA characterizes the same students whose outcomes are measured in NSLDS or in the earnings information from the Department of Treasury described below. Additional information on students' family income and parental education can be gleaned from the FAFSA to help provide context for differences in outcomes for students across institutions. These FAFSA data allow us to report both NSLDS and earnings information for some subgroups based on gender, family income, financial dependency status, and parents' educational levels.

Administrative Earnings Data from Tax Records

To gain insight into the labor market outcomes of individuals attending institutions of higher education, data on federally aided students have been linked to earnings data from administrative tax records maintained by the Department of the Treasury. These linked data are used to produce aggregated and de-identified estimates of institution-level statistics, such as the mean and median of the earnings distribution of federally-aided students in a particular year, along with other metrics described in further detail below. Earnings are defined as the sum of wages and deferred compensation[13] from all W-2 forms received for each individual, plus self-employment earnings from Schedule SE. Importantly, because W-2 forms are filed by employers, the estimates of labor-market outcomes cover the population of employees in the Social Security system and the self-employed individuals who file tax returns. Moreover the earnings information offers national coverage, in contrast to some commonly used information from statewide programs. For example, state unemployment insurance administrative data is limited to the subset of students who work in the same state after leaving college.

Administrative earnings records are used to estimate *inter alia* the following measures of student labor market success for cohorts of federally-aided students beginning their studies in various years:

1. <u>Mean and median earnings among workers</u> (i.e., among those with positive yearly earnings) 6 to 10 years after first enrolling in an institution. This simple measure provides an overall assessment of the degree to which past attendees of an institution are able to find jobs that pay well. Individuals who are identified as currently enrolled (determined by having a federal loan that is in in-school deferment) are excluded from the calculation. To give a sense for the variation in these measures, Figure 3-2 displays variation in institutional median earnings measured 10 years after entry across colleges in each sector. The typical four-year college's median earnings are $40,500 while the corresponding number is $29,800 for two-year and $24,550 for less-than-two-year

---

[13] More precisely, the sum of earnings reported in Box 1 and Box 12 on the W-2 form. This includes both cash and noncash payments. Income related to workers' compensation, small employee achievement awards, adoption assistance, low levels of educational assistance, and small transportation benefits are not included in the earnings calculation (IRS, 2014).

institutions. Control (public, private, or for-profit) also is related to earnings, with for-profits generally having lower earnings.

Figure 3-2: Distribution of Median Earnings 10 Years After Entry

Note: Data are for the 2002 exit cohort.
Source: Department of Education, College Scorecard Data
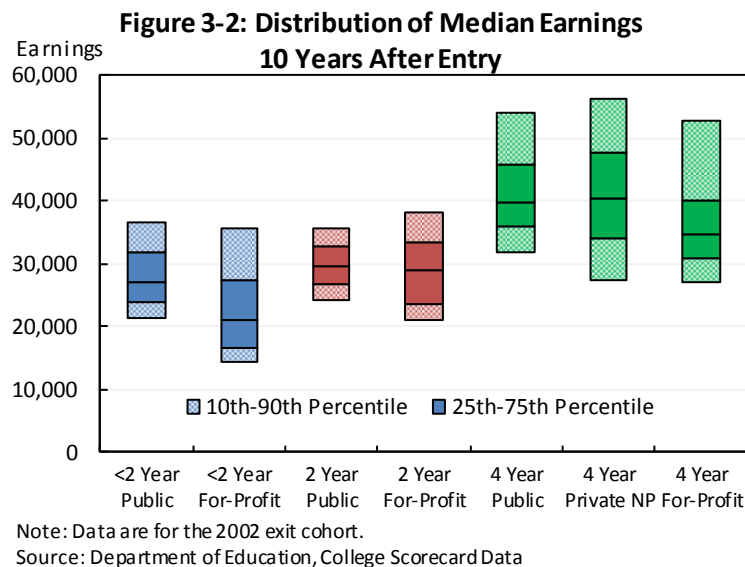
2. The fraction of former students earning over $25,000 (in 2014 constant dollars). This alternate measure of labor market success is designed to measure the extent to which former students find at least a minimal level of employment. In contrast to the estimates of earnings percentiles and average earnings, this measure is based on all non-enrolled individuals (including those with zero annual earnings). The $25,000 threshold was chosen since it approximately corresponds to the median wage of workers age 25 to 34 with a high-school degree only.[14]

3. Percentiles of the earnings distribution for workers. To give a broader sense for the range of earnings outcomes of prior students, for larger institutions the data contain information on the 10th, 25th, 75th, and 90th percentiles of the earnings distribution, in addition to the measures of central tendency shown above.

Important Properties and Limitations of Federal Data

As described above, the data contained in NSLDS, together with earnings information from de-identified tax records, have the potential to significantly expand our understanding of the performance of higher education institutions. There are, however, important aspects of how the data are derived that users of the data should be well aware of to prevent inaccurate or misleading conclusions about institutions and the higher education community at large. The

---

[14] This figure was calculated by using the median earnings from the personal income tables from the 2014 CPS ASEC for high school graduates ages 25-34 for all races. These earnings ($24,835) were reported in 2013 dollars, and were converted to 2014 dollars using the CPI-U (multiplying by 236.712/232.962), giving a final median of $25,234.

Appendix that accompanies this report presents these issues for various new metrics in greater detail, but this section summarizes some of the critical features and limitations of the data that users need to know.

Students Covered

Most of the new metrics of institutional performance described in this report—those based on data in the NSLDS, or matched earnings data for students appearing therein—are based on undergraduate students receiving federal aid. Moreover, institution-specific measures of debt, default, and repayment are based on the subset of students with federal loans. For the purposes of understanding and holding institutions accountable for how well they contribute to the success of students aided by federal dollars, this focus is natural and appropriate.

The lack of data on students who do not receive Title IV aid may result in somewhat biased estimates of collective student outcomes at institutions with low proportions of Title IV aided students.[15] The data thus may not serve as a comprehensive indicator for how well institutions serve all the students they enroll. However, relative to other publicly available data sources that have their own limitations, the data are still likely to be a significant improvement as discussed below. Again, the question of bias depends on the statistic desired—the average outcome of federally-aided or all students—and the differences between those groups within each institution.
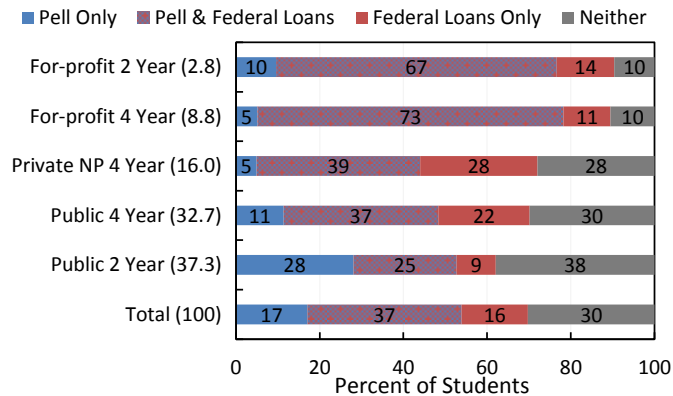
About 70 percent of all graduating postsecondary students receive federal Pell grants and/or federal loans. Overall, Figure 3-3 shows that about 17 percent of enrolled students receive only Pell grants, 16 percent receive only federal loans, and 37 percent receive both. Further, both the fractions of students receiving Title IV aid and the fraction of students with grants, loans or both vary greatly by sector. About 90 percent of all students at for-profit four-year schools and two-year schools receive Title IV aid; in public schools, the analogous Figures are about 70 and 62 percent, respectively. Moreover, in the public two-year (community) college sector, nearly half of Title IV recipients only receive Pell grants (i.e., the median student has no federal debt), whereas about 90 percent of Title IV students at all sectors of four-year colleges and for-profit two-year colleges have federal loans.

Figure 3-3 also depicts the fraction of students who receive federal loans across sectors, using National Postsecondary Student Aid Study (NPSAS) data covering all students (not just federal aid recipients). For measures of borrowing and repayment rates, this is the relevant universe of students for those measures. As can be seen in the Figure, just over half (53 percent) of all students have a federal loan, but there are large differences across sectors. In for-profit institutions, over 80 percent of students borrow with federal loans. Among non-profit institutions, federal borrowing rates are 77 percent at private institutions, 59 percent at public four-year institutions, and about 33 percent at public two-year colleges. Note that the federal role in education financing is even greater than suggested in the Figure, since education benefits like those provided through the GI Bill are not included.

---

[15] Certain measures like repayment or debt to earnings ratios require students to have Title IV aid, so restricting to the Title IV population does not bias estimates for these measures.
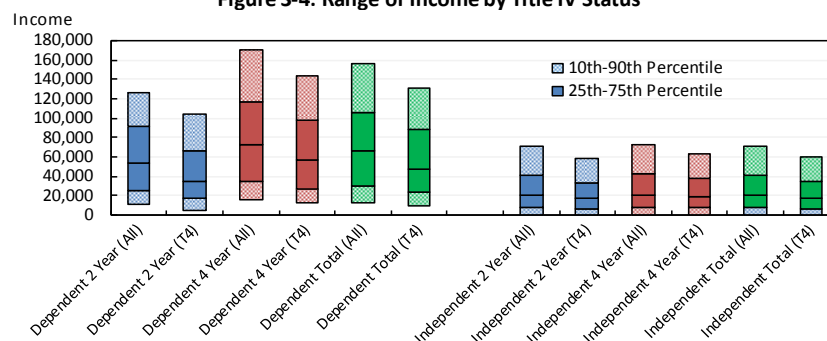
Figure 3-3: Fraction of Students Ever Receiving Pell Grants or Federal Loans by Sector

Legend: ■ Pell Only ■ Pell & Federal Loans ■ Federal Loans Only ■ Neither

Sector	Pell Only	Pell & Federal Loans	Federal Loans Only	Neither
For-profit 2 Year (2.8)	10	67	14	10
For-profit 4 Year (8.8)	5	73	11	10
Private NP 4 Year (16.0)	5	39	28	28
Public 4 Year (32.7)	11	37	22	30
Public 2 Year (37.3)	28	25	9	38
Total (100)	17	37	16	30

Percent of Students

Note: Sector percentage of student population noted in parentheses; data are for students in 2011-2012 who ever received the noted aid type.
Source: National Postsecondary Student Aid Study, 2012

In general, the Title IV population looks reasonably similar to the overall population of a school in terms of student characteristics including ACT and SAT scores, race and ethnicity, age, dependency status, and marital status. This is in part because they represent the majority of students in many sectors. However, there are still some clear differences between federally-aided and non-aided students (2012 NPSAS, CEA calculations). To illustrate an example of these differences, Figure 3-4 below uses data from the 2012 NPSAS and displays the median and range of family incomes for all NPSAS-eligible students (including non-aided students) and for Title IV students by school sector. Although Title IV students are clearly lower-income than the general student body (with median family income of $47,800 vs. $65,500 overall for dependent students and $17,000 vs. $20,000 for independent students), the range of incomes overlaps considerably.

Figure 3-4: Range of Income by Title IV Status

Income

Legend: ☐ 10th-90th Percentile ■ 25th-75th Percentile

Categories: Dependent 2 Year (All), Dependent 2 Year (T4), Dependent 4 Year (All), Dependent 4 Year (T4), Dependent Total (All), Dependent Total (T4), Independent 2 Year (All), Independent 2 Year (T4), Independent 4 Year (All), Independent 4 Year (T4), Independent Total (All), Independent Total (T4)

Note: Dependent income includes parents' income while independent income includes student and spouse's income. Data are for students in 2011-2012. Title IV status is based on a student ever receiving Title IV aid.
Source: National Postsecondary Student Aid Survey, 2012

27

Institutions Covered and the Level of Aggregation of Information

The data presented in this report cover the universe of all institutions of higher education that are active participants in the Title IV federal aid program, are located in the 50 states and the District of Columbia,[16] and both have a positive number of degree-seeking undergraduates and awarded some undergraduate credentials in the 2012-2013 school year. The IPEDS definition of institution is used as the basic level for reporting data, though for many analyses in this report a higher level of aggregation (i.e., six-digit OPEID) is used due to data limitations. Postsecondary institutions are complex organizations, often comprising many separate campuses, including online programs, and sometimes operating under different names entirely. Each organization chooses which of its several branches it would like to report information for separately, as opposed to jointly with the 'main campus,' generally aligning with how the institution represents its collection of sub-entities to the public. There are 575 main campuses that have at least one or more additional branch campuses in the data in the universe described above, and IPEDS assigns each a unique identification number called a unit ID.

These institutions vary dramatically in size across sectors. For example, although public four-year schools educate over a third of all undergraduates, they account for less than 10 percent of institutions. A similar trend is apparent for public two-year schools. On the other hand, for-profit less-than-two-year schools educate less than 5 percent of students but account for over a third of all schools.

Table 3-1: Distribution of Students and Schools by Sector		
	Fraction of Students	Fraction of Schools
Public <2 Year	8.2%	8.9%
Private Non-Profit <2 Year	0.2%	2.9%
Private For-Profit <2 Year	3.6%	35.7%
Public 2 Year	30.2%	11.3%
Private Non-Profit 2 Year	0.8%	2.3%
Private For-Profit 2 Year	2.3%	8.5%
Public 4 Year	34.9%	8.3%
Private Non-Profit 4 Year	15.3%	17.9%
Private For-Profit 4 Year	4.4%	4.3%
Source: Department of Education College Scorecard Data		

---

[16] Note the underlying data contain information on institutions in the U.S. territories as well, but these are omitted from the report primarily because they often have outlier outcomes for earnings measures.

Some complications in comparing institutions arise when considering branch locations. The IPEDS data and NSLDS data have different reporting requirements for branch locations, and institutions also have latitude in how they aggregate campuses and programs in reporting.[17]

About two-thirds of institutions, collectively enrolling 83 percent of students, have only one main campus identifier (or six-digit OPEID) assigned and are thus not affected by this issue. Roughly 40 institutions have more than 10 branch campuses (noted as separate eight-digit OPEIDs), and some have many more. For example, ITT Technical Institute has 140 campuses (unit IDs), and Strayer University has 100. In addition to these large for-profit systems, some large public university systems can also have many campuses, yet report at a single, six-digit OPEID. For example, the Pennsylvania State University has 23 branches. Because outcomes at branch campuses may vary greatly, the NSLDS data may be inaccurate for any given branch, and advanced users of the data may wish to either exclude such observations from their analyses or use statistical modeling techniques to impute branch specific outcomes.[18]

While some institutions assign only one six-digit OPEID and differentiate at the eight-digit level, it is worthwhile to note inconsistent organization at the six-digit OPEID level, as some large public and for-profit systems assign different OPEIDs for each campus. For example, the California State system has twelve different six-digit OPEIDs assigned to it, and the University of Wisconsin system has thirteen (this differs from the Penn State example above). Some for-profit companies also have many six-digit OPEIDs; Aveda and Kaplan have 29 and 20 six-digit OPEIDs assigned, respectively.

---

[17] The Office of Federal Student Aid (FSA) recognizes the separate branch 'locations' for all Title IV-eligible institutions that have a Program Participation Agreement (PPA)—the formal document establishing their eligibility to participate in Title IV programs—with the Office of Postsecondary Education (OPE), and assigns each location an 8-digit OPE identification number (OPEID). Related institutions all share a common 6-digit root tied to the "main campus" that is the entity that enters into the PPA with FSA. Institutions may choose to list additional locations as a separate entity (or branch campus) in IPEDS so long as the institution is permanent, is physically removed from the parent (main campus) institution, and offers organized programs of study (e.g., rather than a location for offering courses only). Aside from closures, mergers, and new schools, institutions of higher education may revise their reporting structures in IPEDS. For instance, the University of Phoenix requested to change its reporting beginning in 2014-15 from campus-level to state-level. Therefore, over time, OPEIDs for the institution are not consistent. While NSLDS is increasingly moving towards reporting student enrollments and aid-receipt at the 8-digit OPEID level, this is not yet universal and many institutions report information on where students enroll or receive aid only at the 6-digit OPEID level (since the main campus is often where aid is managed for all branches of an institution). At present, therefore, data derived from the NSLDS and earnings measures that take the universe of Title IV students at each institution as a base can be reliably calculated only at the 6-digit OPEID level, a limitation shared by the familiar cohort default rate measure. Thus, IPEDS institutions sharing a common 6-digit OPEID are all assigned the (student-weighted) average outcome or median outcome for students across all branches of the institution for NSLDS or tax-data derived measures.

[18] The documentation to the data described in this report includes a comprehensive crosswalk between the unit IDs used in IPEDS and the OPEIDs used by FSA to administer federal aid programs.

Cohort Definitions Are Imperfect and Vary for Different Metrics

The data contain diverse measures of institutional performance constructed both with an eye towards the type of information that would be most useful to prospective students, as well as towards how the measures might promote accountability for institutions. The measures require different definitions of cohorts. Users of the data should be aware of this, particularly when constructing analyses of the relationship between different measures. Moreover, reporting inaccuracies in some data elements used for cohort definitions are also important.

Data for both earnings and NSLDS-constructed completion rates are based on cohorts of students who are estimated to have begun their studies in the same "*entry year*"—measured in terms of "award years," which run from July 1st to June 30th (e.g., award year 2002 is from July 1, 2001 to June 30, 2002). While this construction is similar to the completion cohorts reported in IPEDS, limitations in the data create measurement error, the extent of which is difficult to assess. The NSLDS data do not directly record the date when students first enter an institution. That date is estimated based on a combination of when the student is first observed receiving federal aid at an institution, and the student's self-reported grade level on the FAFSA form associated with that record. In particular, if students report they are entering their second undergraduate year on the FAFSA associated with their first receipt of aid at a university, they are assigned to an entry year one year prior to when we observe them first receiving aid. For students reporting that they are entering their third through fifth undergraduate year, they are assigned an entry year two years prior. The adjustment is capped at two years, since it appears that a non-trivial portion of respondents who report entering their fourth or fifth undergraduate years are misreporting their secondary school grade level on the FAFSA form.[19] Limiting the adjustment in this manner avoids (though perhaps overly so) penalizing institutions by improperly assigning students to an earlier cohort, where the student would be less likely to be counted as a success in the institution's completion rate calculation. In sum, the NSLDS completion and earnings cohort measures are thus based on the universe of Title IV students in a given entry cohort at each institution. IPEDS completion measures are based on the set of first-time full-time students in a given entry cohort at each institution. While in principle the cohorts contain only undergraduate students, in practice it appears that inaccurate reporting may lead to graduate students being included in the cohorts in some cases. Students are included in the cohort if they receive either a Pell Grant or an undergraduate federal loan, where an "undergraduate loan" is identified as those where an institution does not report the academic level of student as a graduate student on the loan record. There seems to be some evidence of misreporting by academic level, for example, as we find undergraduate students enrolled in some schools known to enroll only graduate students. Another way in which this might occur is if students receive an undergraduate loan at an institution, but then begin a graduate program at the same institution with no changes recorded for their academic level in NSLDS. Both types of error appear limited in scale, but might affect results for some institutions. To a certain extent, we address this type of error by eliminating institutions that report no degree-seeking undergraduates or undergraduate awards in IPEDS.

---

[19] For example, a student who is first aided in 2003 who reports entering his second undergraduate year would be assigned to the 2002 entry cohort. An exception to this general rule is that students who are observed transferring from a previous institutions are assigned to the entry cohort when they are observed first-aided at the institution.

Student debt measures are based on cohorts of federal loan borrowers who exit from institutions in a given year. That is, cumulative student debt is measured both for students who withdraw and those who graduate from an institution in a given year, again based on award years. The data are provided as a single metric and then disaggregated by completion status.

Finally, repayment rates are based on the set of federal loan borrowers who enter repayment in a given repayment year, corresponding to the federal fiscal year; so the 2011 repayment cohort is based on students entering repayment from October 1, 2010 to September 30, 2011. Since students who graduate may not immediately enter repayment either due to their 6-month grace period or to a loan deferment granted, for example, for economic hardship or upon entering graduate school, many students are likely to enter repayment in a different year than when they exit the institution.

Since different metrics are based on different groups of students and organized by different event dates, users of the data should exercise caution when linking data for different metrics for the same institution. For example, an analyst may be interested in how changes in completion rates affect the likelihood of loan repayment across a set of institutions. This analysis is complicated because neither NSLDS nor IPEDS measures of completion are presented only for borrowers (the relevant population for the repayment measure), and because the set of students in a given entry year cohort may enter repayment across several different repayment years. To partially facilitate linking measures for such analyses, the completion and earnings data contain the median dates at which individuals in each entry cohort exited the institution (relevant for debt) and entered repayment (relevant for repayment rates).

Using Institutional Data to Explore Sector-level and Student Outcomes

While the data presented here are organized by institution, they can in some cases be used to inform student outcomes at various levels of aggregation such as for a particular institution type or sector, or student subgroup (e.g., low-income students). For many institutional outcome measures reported, including completion or repayment rates, the data include the number of students in the institution and the subgroup upon which each measure is based (i.e., the number in the denominator of the rate) so that a weighted student level average can be completed by aggregating outcomes across any relevant set of institutions.

There are, however, some limitations to subgroup calculations that are important to understand. First, as mentioned above, the data are only for Title IV aid recipients. This makes it difficult to calculate what fraction of the relevant subgroup is represented by the data because IPEDS and other data sources do not contain the total of all Title IV students or all undergraduate students in a given entry or exit cohort. Moreover, some data elements are suppressed for institutions or subgroups with few observations, and so a small fraction of students may not be represented in the data. Additionally, to reduce the influence of outlier observations and provide students with more reliable information, many measures of performance such as cumulative debt levels and earnings information are presented for the median (rather than mean) student in a cohort. Medians cannot be aggregated in the same way as means to calculate, for example, median

earnings for students in a particular sector. In some cases, both means and medians are presented, so users can select the measure that best suits their purpose.

Additional Notes on the Earnings Measures

Beyond the broad data issues described above, users of the earnings data should be aware of several details that give context to the interpretation of particular student outcome metrics. As described above, aggregate earnings statistics are presented for students measured at various points in time following the year they enter an institution. Data users should be aware of the following important aspects of these data:

1) <u>Earnings are estimated for undergraduate Title IV recipients only</u>, because we match student-level data in NSLDS to administrative earnings records. The share of each institution's entering class represented by Title IV students can vary substantially due to both differences in family income of students attending those institutions and to state and institutional aid policies. In institutions where this share is low, results may be less representative of the entire student body. However, as noted above, over 70 percent of all graduating postsecondary students have ever received either federal Pell grants or federal loans under the Title IV program.

2) <u>Mean and median earnings, as well as percentiles of the earnings distribution, are presented for non-enrolled workers</u> (where "worker" is defined as an individual with positive earnings in the calendar year). Enrolled individuals are omitted from the calculation so as to avoid having earnings appear low for institutions where a high fraction of students go on to graduate studies (though some enrolled individuals are in different undergraduate institutions). These students are identified as enrolled if they have an undergraduate loan that is in in-school deferment status for at least 30 days during the measurement year. A limitation of this methodology is that there is no way of identifying whether former Title IV recipients who received only Pell grant aid are enrolled. Supplemental analyses performed by the Department of the Treasury compared estimates of enrollment status of students based on the deferral status of their loans to estimates based on the existence of 1098T tax forms filed by institutions to allow students to deduct tuition expenses from their taxable incomes. While both measures lead to similar estimates of the fraction of students enrolled, different students are identified by each measure. Although further work is required to fully understand the measurement gaps, preliminary results suggest we may be mistakenly including some enrolled students in the earnings measure. This could lead to an underestimation of mean and median earnings of non-enrolled students at institutions where students significantly reduce their work hours while enrolled.

3) <u>Earnings are measured for all Title IV students who attended an institution, regardless of completion status.</u> Thus, variation in earnings across institutions will partially reflect differences in completion rates (to the extent completion affects earnings at an institution). This differs from other commonly reported earnings measures that often present earnings information for graduates only, ignoring the outcomes of the often

numerous students who fail to complete their degree. Moreover, PayScale data both relies on voluntarily reported data for graduates, sometimes based on relatively few reporters, and ignores the earnings of students who continue their studies at a graduate level, potentially biasing down earnings measures for schools that send many graduates on to further study Many public state systems rely on state unemployment insurance data, which are limited by the inability to track students across states. It should also be noted that measured earnings might reflect the contributions of a student's subsequent education. For example, if a student transfers to a different institution, or attends graduate school, and is working and not enrolled at the point of measurement, their earnings are captured in the cohort of the original institution. In this sense, the measure captures all the ways in which an institution might contribute to its students' labor market success (i.e., supporting them through graduation, and helping them transfer or be accepted to graduate school).

4) <u>The data include information on the fraction of individuals who did not work for pay</u> among those who are not currently enrolled. This is based on information about the number of individuals with no reported earnings over the course of the full year. Data users should be careful in interpreting this as a measure of unemployment, meaning the fraction of workers in the labor force (actively searching for a job) who are unable to find employment. Based on the Current Population Survey, for example, in 2014 only about one-fourth (25.6 percent) of non-employed individuals not currently enrolled, with at least some college experience, and between the ages of 25 and 34 say they are currently looking for work (i.e., are unemployed).[20] Of course, the fraction of those not working who are likely to be "unemployed" is likely to differ across institutions. For example, at institutions that specialize in vocational training offerings, it is likely that a significant share of non-employment does in fact represent unemployment, given the goals of students enrolling in such programs.

Additional Notes on NSLDS Completion and Transfer Rate Measures

This report relies primarily on the well-known IPEDS completion measures. As explained above, the data also include preliminary, NSLDS-derived, institution-specific measures of completion and transfer rates that appear very promising relative to currently available measures and are likely to improve as institutional reporting becomes more accurate. Since NSLDS is not designed to measure these outcomes, however, it is not surprising that there are some limitations to using the data to measure completion and transfers. These data are currently not appropriate for consumer information purposes, pending further quality reviews. Researchers should therefore be careful in their use and especially cognizant of the following limitations:

1) <u>While many institutions appear to report completion or withdrawal outcomes very accurately, many struggle to do so for their students who do not take out loans.</u> This is primarily attributable to the fact that NSLDS is an administrative financial aid database, the main purpose of which is tracking such information to determine when students enter

---

[20] Based on CEA calculations using the combined monthly files in the 2014 Current Population Survey.

repayment. This decision does not depend on whether students withdraw or graduate and is irrelevant for students who do not borrow to attend school. As a result, reporting for Pell-only recipients is inconsistent, leading to low estimated completion rates in many schools with high fractions of Pell-only recipients. A comparison with state administrative data in one state shows that over 80 percent of Pell-only completions were not reported for the 2007 cohort, though that fraction dropped below 50 percent in the next cohort. In particular, measured completion rates are particularly low for community colleges at which many students receive only Pell grants, and also appear low for some elite institutions that have adopted "no-loan" financial aid policies for students below some family income threshold (so Title IV recipients are predominantly Pell only recipients). More generally, some schools appear not to reliably report completion outcomes for any students (those receiving loans and/or Pell grants) and therefore have (implausibly) low estimated completion rates.[21]

2) <u>The data does not directly report start dates for students</u>. Rather, start dates are estimated from the dates we observe students first receiving federal aid at an institution in conjunction with their self-reported grade level at that time. As a result, there may be errors in assigning students to cohorts. Based on attempts to validate the completion rates using administrative data from a state higher education system, it appears that this method assigns students to the correct cohort about 70 percent of the time, and to an adjacent year's cohort another 10 to 12 percent of the time. Larger errors tend to assign students to more recent cohorts, which may have the effect of artificially increasing cohort completion rates by extending the measurement window. Since relatively few completions occur after 150 percent of the program length, this should not have a large impact on measured rates.[22] Validation analyses using state administrative data suggested fewer than five percent of students overall were wrongly assigned to a more recent cohort.

3) <u>Finally, the data in NSLDS on enrollment intensity and transfer status are both of poor quality</u> for Pell-only students prior to 2012. Because of this, the data do not support reporting completion rates disaggregated by full-time and part-time status, or first-time and not-first-time status. Moreover, since transfers can only be identified if the student receives Title IV aid at the transfer-in institution, NSLDS cannot reliably identify all transfer students.

---

[21] NSLDS data were compared with data from other sources, such as data from the National Student Clearinghouse (NSC), the State Council of Higher Education for Virginia (SCHEV), and the Beginning Postsecondary Survey (BPS). These data were used to conduct national-, sector-, and student-level comparative analyses, using NSC supplemented data where appropriate.

[22] Overall, an additional 3 percent of students complete between 150 percent and 200 percent time frames. This varies by sector, where two- and less-than-two year public schools, two-year non-profits, and four-year four-profits see more completions between those time frames.

IV. Methods for Assessing the Causal Effect of Institutions on Student Outcomes

There are a variety of techniques that have been developed to measure the causal effect of institutions on student outcomes. This is an active area of methodological development and research. This section provides an overview of the various approaches to estimating institution quality, and discusses the strengths and weaknesses of each approach. It also presents exploratory estimates of institution quality that illustrate how differences in methodology can lead to quite different estimates of college quality.

An Overview of Methods

There is a deep and active academic literature in economics and education research on how to best measure schools' causal impact on student outcomes, though until recently it has largely focused on K-12 education. This section presents an overview of the benefits and drawbacks of different statistical approaches to measuring the causal effect of postsecondary institutions and discusses details of exploratory estimates of institutional quality based on federal data. An in-depth treatment of these estimation issues is beyond the scope of this report, but several excellent discussions are available including Hoxby (2015), Chetty, Friedman, and Rockoff (2014), Rothstein (2010), and Kane and Staiger (2008).

Consider the following model that relates student outcomes Y_{is} to their characteristics prior to beginning their studies and school quality

$$Y_{is} = X_{is}\beta + \mu_s + \epsilon_{is} \qquad (1)$$

where X_{is} is a vector of observed student characteristics including family income, gender, etc. that are associated with students' outcomes; μ_s is a vector of school fixed effects (i.e., coefficients associated with a vector of indicator variables equal to one if student *i* attended school *s)*; and ϵ_{is} represents all unobserved determinants of outcomes (e.g., academic background, career goals, work ethic, etc.). The parameters of interest in this framework are μ_s—the subscript *s* is meant to underscore that there is one coefficient associated with each school—which represent the causal effect of attending school *s* on student outcomes relative to attending some reference institution, or in other words, the institution's "quality."[23]

There are several challenges involved in accurately estimating schools' causal contributions to student success along a particular dimension. Most importantly there may be omitted variables— student characteristics that are important determinants of outcomes—that might hinder attempts to statistically control for differences in the types of students that different institutions enroll. For example, an important limitation of federal data sources is a lack of information on

---

[23] That is, if *Y* is annual earnings, then μ_s has the interpretation of the earnings that a student would earn by attending institution *s* relative to what he would earn by attending some reference institution, conditional on the vector of characteristics X.

individual students' academic preparation, such as high-school GPA or college admissions test scores (e.g., SAT or ACT scores). Since academic preparation is likely to be both related to the outcome of interest and to college quality, omitting this variable may bias estimates of college quality. Of course, the impact of omitting any particular variable on the accuracy of college quality estimates depends on context. Controlling for differences in academic preparation may be less important, for example, in estimating quality among less- or non-selective schools where differences in the academic background of students across schools may be less pronounced.

A related challenge is that students may enroll in colleges of differing quality depending on the observed and unobserved characteristics of schools and student preferences, and these determinants of enrollment choices may affect outcomes. Hoxby (2015) explains this selection can be either "vertical," when institutions are more selective in their admissions and the correlation between the selection criteria and outcomes is the source of bias, or "horizontal," when schools may not be selective at all but may nonetheless enroll students with different "potential outcomes" due, for example, to geographic differences in the strength of the labor market or proximity to better high-schools and thus better prepared students. Depending on the estimation strategy chosen, both omitted variables and selection effects can lead to misleading estimates of school quality, where these biases may potentially lead to overestimating or underestimating school quality.

Much of the methodology in K-12 education literature focuses on isolating the causal impact of schools and teachers on student outcomes. It is worth emphasizing that accurately measuring the impact of attending an institution may be more empirically challenging than in K-12 education for a number of reasons. First, it is often impossible to observe a particular outcome (e.g., earnings or repayment) for students prior to their entry into college, whereas, at least in elementary and secondary school where test scores are a common outcome, it is common to have standardized tests in a subject measured at the end of each grade. The lack of a "pre-test" to include in X_{is} rules out research designs (e.g., individual fixed effects models) that use the change in an outcome for the same individuals to better control for hard to observe but "time-invariant" (i.e., not changing over time) differences in the characteristics of individuals in one school versus another.[24] And second, it seems likely that student sorting into colleges based on quality is likely to be more pronounced than student sorting to K-12 schools or teachers given the more constrained choices offered in the K-12 level for most students. As a result, the bias in raw outcome measures may be greater in magnitude, leaving more work for the statistical model to isolate quality.

Below, several methods for measuring college quality are discussed, along with the strengths and limitations of each method.

---

[24] While some students—especially older students—may have labor market experience prior to entering college, it can be difficult to judge whether the time path of their earnings prior to entry—which in some cases might represent part-time earnings while enrolled in high-school—are a good counterfactual to what their earnings would have been as they aged (e.g. and graduated high school) if they did not attend a college. For a measure such as a repayment rate, no pre-college data are likely to exist.

Aggregate (School-Level) Regression Adjustment

A commonly used method for estimating college quality relies on aggregate institution-level data. In particular, a school's causal effect on student outcomes is measured by regressing school-level average outcomes ($\bar{Y}_S$) on school-level characteristics, including the average characteristics of students at the school ($\bar{X}_S$), and designating the residual from this equation the school's quality, where quality is implicitly measured as the component of a school's average earnings that is not explained by the school's characteristics (including the average characteristics of its students). The logic behind this approach is intuitive: the regression model estimates the "predicted" average student outcome of each college based on the relationship between these outcomes and school level inputs, such as the average SAT scores of students, the fraction who are low-income, the fraction female, etc. The difference between the school's actual and predicted outcomes, is used as the measure of quality—in other words, quality is the portion of the institution's students' success that is not "explained" by the students' characteristics. The rankings published by *U.S. News and World Report* and *Washington Monthly* magazines both include such a measure for completion rates, albeit weighed lightly, and researchers have also used these types of measures for both earnings and completion (Rothwell and Kulkarni, 2015; Kelchen and Harris, 2012).

An important consideration in implementing any of the approaches discussed in this section, is the choice of which student and institution characteristics ($\bar{X}_S$) to include in the model. In principle, to measure the causal impact that institutions have on student outcomes, it is important to control for all predetermined (measured prior to a student's enrollment) student characteristics that might affect outcomes and are likely to differ across institutions. The most important set of variables to include depend on context, and this report generally considers only those relevant for comparing performance across the full range of post-secondary institutions. Including institution level data may also be appropriate, but doing so might risk masking some of the institutional quality effects we are attempting to measure by conditioning on mechanisms that colleges use to produce good outcomes for their students. For example, inclusion of per-student expenditures as a control variable might absorb the part of institutional quality related to the resources institutions' devote to students. The measure of quality that results could still be relevant—for example if the relevant question has to do with how efficiently institutions produce good outcomes for their students—but may be misleading to students making college choices. In general, controlling for institution level determinants of quality (spending, teacher characteristics, class-size, etc.) may similarly lead to misleading information about quality differences.

While the aggregate model is attractive because data on aggregate institutional outcomes and student characteristics are more readily available, it is unlikely to yield accurate measures of differences in college quality across a broad and heterogeneous set of institutions.[25] This method relies on the assumption that student characteristics are uncorrelated with the quality of the institution in which they enroll, or in statistical parlance, treats μ_s as a "random effect." If this

---

[25] As noted above, the method may be appropriate in a more homogenous subset of schools, or where there is less selection of particular types of students to higher or lower quality institutions.

assumption is false and, for example, students from higher income families sort into higher-quality schools and family income independently affects completion rates, then the estimated relationship between completion rates and family income will reflect both this student-sorting selection effect and the higher quality effects of the schools with richer students. The aggregate-level regression model treats all of this correlation as attributable to family income differences across schools, however, and so will adjust away this part of the quality differences across schools. Indeed, in a recent working paper Altonji and Mansfield (2014) argue that if a sufficiently rich set of information is included in the vector X_{is}, then the residualized outcomes (i.e. the regression residuals used for the school quality estimates) from the aggregate regression will capture only the component of quality that is unknown to students and their families when they select an institution to attend.
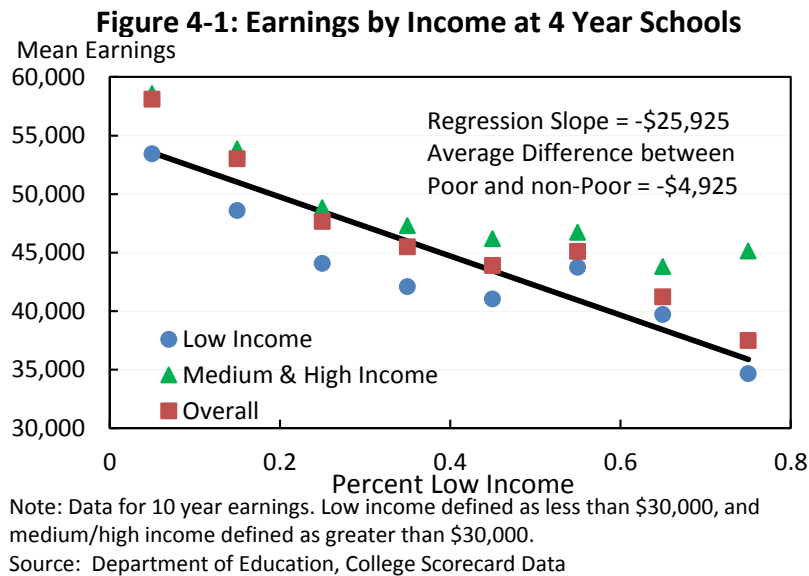
A simple illustration may help convey the intuition for how aggregate regressions can mislead. The solid square markers in Figure 4-1 depict the average of institutional mean earnings 10 years after entry, across eight x-axis "bins" of the share of the institutions' students who come from families with incomes below $30,000, for four-year schools. For example, the first square on the left of the Figure represents the average earnings for Title IV students who attend institutions where between zero and 10 percent of students come from low-income families, and the next square to the right depicts the same information for students in institutions with between 10 and 20 percent of students coming from low-income families, etc. The estimated regression line of institutional average earnings for the fraction of students who are poor suggests that average post-college earnings are about $26,000 lower for low-income students than for students who are not low-income. Accordingly, the aggregate regression approach would judge an institution that was comprised of all poor students to have the same quality as an institution with no poor students but with $26,000 higher earnings.

The triangular and circular markers in the Figure, however, show the average incomes for these two different groups of students within institutions in each range of low-income student enrollment share. The Figure shows that within-school differences are much smaller than across-school differences, only $5,000, about one-fifth the magnitude implied by the aggregate regression line, and do not vary much across institutions. It is clear that students in schools with more low-income students have worse outcomes for all students, not just poor students, suggesting that the observed relationship between aggregate earnings and the share of low-income students may reflect more than the direct impact of students' family incomes on earnings.[26] This is suggestive evidence that school quality is correlated with the family income of enrolled students and is consistent with the argument that aggregate regressions may over control for quality (i.e., the regression line in the Figure is drawn too steeply).[27]

---

[26] Other quality measures may be related to student backgrounds in similar ways as well. However, not all are related in every context. For instance, student characteristics exhibit a negligible relationship with the gainful employment metric (debt to earnings ratio) among applicable programs; this is much less the case when comparing a broader set of outcomes across the full spectrum of higher education institutions.

[27] Another possibility is that the observed difference in the across-school and within-school correlation between family income and earnings is driven by an omitted factor that is positively correlated with both family income and

Figure 4-1: Earnings by Income at 4 Year Schools

Note: Data for 10 year earnings. Low income defined as less than $30,000, and medium/high income defined as greater than $30,000.
Source: Department of Education, College Scorecard Data

School-Level Matching (Peer Comparison) Approaches

An approach that is closely related to measuring an institution's causal impact with an aggregate-level regression is to group colleges into sets of "peer" institutions and then to compare the raw outcomes of each institution within the group. In this framework, the peers of an institution are typically selected based on the affinity of aggregated student-level characteristics and some institution-level characteristics. The logic of this approach is essentially the same as in the aggregate regression approach: The estimate of causal impact for each institution is based on a comparison of its actual outcome to a "predicted" outcome, but the predicted outcome is derived based on the average outcome for the set of peer institutions rather than the regression model. This may have some advantages over the regression approach described above: a) it can be more transparent to a non-technical audience, in that it is based only on the actual student outcomes of the institution and the peer group (though in practice these comparisons are also complex, since sophisticated matching algorithms just move the statistical modeling to the selection of peer groups rather than the adjustment of outcomes); b) to the extent groupings reflect institutions' senses of their own peers, this type of information may be useful in allowing them to benchmark their performance; and c) as a technical matter, to the extent that the relationship between student and institution characteristics may be highly non-linear, this approach can be less sensitive to errors in specification in the regression model.

On the other hand, peer group methodologies may suffer from the same drawbacks as the aggregate regression model, and there may be other sources of error as well. If peer groups are formed by partitioning institutions into a set of groups, and each institution is compared to its group's average performance, then estimates of causal impact can swing dramatically for institutions that might be close to the group boundaries. Alternatively, if peer groups were created by splitting schools into selectivity tiers based on their students' average SAT scores,
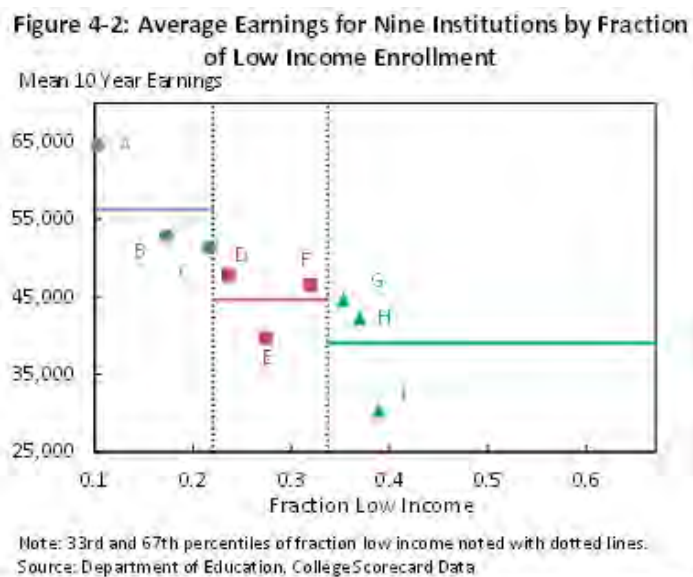
---

post-college earnings. For example, if both richer and poorer students at wealthier schools tend to have stronger academic preparation, this could account for the "excess" correlation in family income and earnings across schools.

some schools near the cutoff points might be evaluated differently despite having very similar student composition and student outcomes. For example, the school with highest student SAT scores in the lower selectivity group is likely to be evaluated more favorably than the school with the lowest (but similar) SAT score in the higher selectivity group even with identical outcomes because of the difference in peer groups. Similarly, if institutions were compared within the same geographic region, the rating of institutions near these boundaries could be distorted.

To illustrate this point, Figure 4-2 depicts the earnings outcomes of nine institutions and considers a case in which institutional peer groups are formed by grouping institutions into three groups based on the share of low-income Title IV students they enroll, indicated by the vertical dotted lines. In the Figure, the mean outcome within each group is indicated by the horizontal line. To see the potential for misleading results near the boundary, consider how this approach would assess institutions C and D. College C has higher earnings than D and only a slightly smaller fraction of low-income students. However, since it is just below the low-income enrollment share delineating the first and second peer groups, its earnings are compared to a higher peer group mean, suggesting it provides worse outcomes for its students than D, which looks good as it is compared to a lower peer average.

Another limitation of this approach is that with fairly coarse groupings, the approach might not sufficiently match student characteristics between schools, allowing differences within groups to be influenced by fairly large differences in student characteristics across institutions. For example, College B has only slightly higher earnings than C, but C enrolls almost double the number of low-income students—a difference that is ignored by this approach since both schools are in the same group.

Figure 4-2: Average Earnings for Nine Institutions by Fraction of Low Income Enrollment

Note: 33rd and 67th percentiles of fraction low income noted with dotted lines.
Source: Department of Education, College Scorecard Data

This example also shows that, similar to the aggregate regression approach, comparing school performance relative to their peer group means may obscure too much information about school quality. Despite having significantly higher earnings, College B would be rated worse than College

H since they are compared to very different peer institutions. If, as suggested above, higher family income students tend to sort into higher quality institutions, then this may not be appropriate.

Of course, there are many different ways of implementing a peer-group methodology. For example, finer approaches exist that select a different set of peer institutions for every institution. The website College Results Online, for example, separately selects peer institutions for each college by identifying a set of institutions that is most similar based on SAT/ACT test scores, the percentage of students receiving Pell grants, the type of control (public vs. private), total full-time enrollment, and a host of other factors (College Results Online, 2015). This approach helps in the creation of a more homogenous set of peer institutions, but may sacrifice some simplicity and transparency in the process.

Individual Level Model of Causal Impact

To control for the important issue of student sorting based on quality, economists tend to model μ_s as a vector of fixed effects, with the associated coefficients representing the institutions' quality, or causal effect on the outcome Y_{is}. The innovation of this approach over those discussed above is that the within-school covariation between student characteristics and outcomes are used to identify β, and thus controlling for those characteristics does not absorb across school differences in quality that may be related to average differences in X_{is}. For concreteness, in the context of the discussion of Figure 4-2 above (and ignoring any role of other characteristics), a school with all poor students would be judged equivalent in quality to a school with no low-income students and $5,000 higher earnings—the average difference in earnings between students from low-income and not low-income families within schools (rather than $26,000 higher earnings using the aggregate level approach).

In this "institution fixed-effect" approach, if students select institutions based on omitted characteristics that are related to outcomes conditional on X_{is}, then estimates of the causal impact may still be substantially biased (as would be also the case with the aggregated institution-level data-based methods described above). The extent of this problem depends on the particulars of the data available. In the federal data used in this report, rich information on students' family income prior to school entry is available as a control, as is information about the student's gender, age, veteran status, and parents' education levels. Borrowing from a strategy employed by Dale and Krueger (2002, 2011), data on the number of institutions to which students send their FAFSA, and the average SAT (equivalent) score of those institutions (as reported to IPEDS) are also used as controls. These variables may help proxy for the student's academic preparation and ambition—an important addition since federal data do not contain individual-level information on a student's academic ability such as standardized test scores or high school GPA. The key question for whether this method is able to deliver unbiased estimates of institutions' causal impacts on student success is whether unobserved factors (such as individual level academic background) both affect the outcome of interest (earnings or completion rates, for example) and are correlated with other variables included in the model.

Comparing the Use of Individual-Level versus Aggregate-Level Approaches to Measuring the Causal Effect of Institutions on Student Earnings

This section presents institutional quality measures based on post-enrollment earnings estimated by the aggregate and individual level regression approaches, and compares them to raw average student outcomes. It should be stressed that the estimates presented here are exploratory, and intended to facilitate a discussion of methodological issues and further work necessary to develop accurate institutional quality measures. They are not intended to be viewed uncritically as accurate estimates of institutions' causal effect on earnings.

To estimate college quality in the institutional fixed-effect regressions, attention is restricted to annual earnings measured 10 years after students first enter the institution, and the sample is limited to students with positive earnings who are not currently enrolled in school. These students entered their institutions in either the 2001 or 2002 award years, and their earnings are measured in the 2011 or 2012 tax years. All of the regression analyses are estimated separately by institution level (four-year, two-year, or less than two-year schools) based on the level of the predominant degree awarded by the institution in 2013.

Several variations of equation (1) are estimated, each adding richer sets of individual-level control variables to X_{is}, taken from information on the FAFSA forms associated with students' first records of federal aid receipt. The first specification controls primarily for socio-economic status variables, including student's financial dependency status, the log of their family income interacted with independent status, and indicator variables for parents' highest level of education completed. In the second specification demographic variables are added, including the individual's age and age-squared measured when they first enter the program, gender, marital status, veteran status, and whether they have children. A third specification is similar to the second, but controls for family income in a more flexible way by replacing the log of family income variables with five indicator variables corresponding to ranges of family income, again fully interacted with independent status (the categories are $0-$30,000, $30,001-$48,000, $48,001-$75,000, $75,000-$110,000, and greater than $110,000, all based on income bands used to report net price differences in IPEDS). Finally, the fourth specification adds a set of indicators that describe the number of schools a student sends a FAFSA form to (dummies for 1, 2, 3, 4, and 5 or more), and eight dummy variables for the average SAT equivalent score for the schools to which the FAFSA was sent.[28]
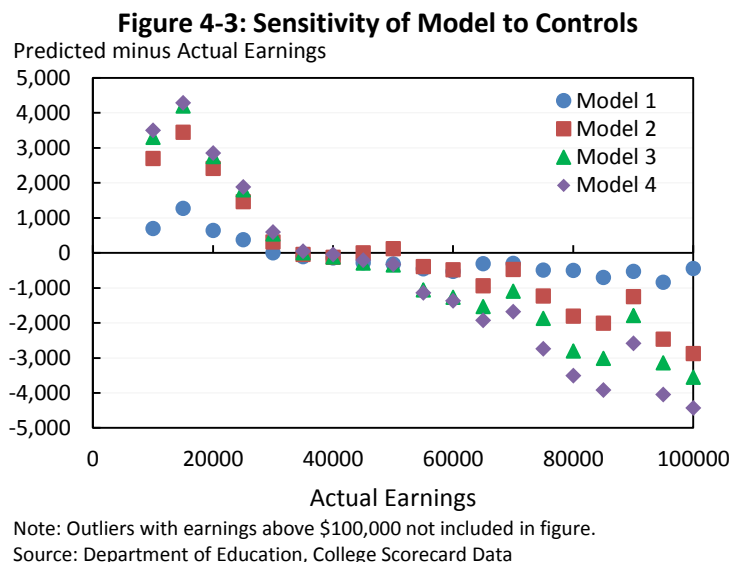
The models described above do not necessarily reflect the frontier of what might be done using federal data. Rather, they control for a simple but important set of demographic information that any approach would likely adopt, and that can be used in both individual-level and aggregate-level models to highlight how results depend on geography. Federal data lack important individual level controls for race and ethnicity and academic background (e.g., individual level SAT scores), so they are not included. Controlling for differences in local labor market strength

---

[28] Data from IPEDS are used to form average SAT score equivalents using reported ranges for ACT and SAT scores. The eight indicators correspond to whether this score lies in the following categories: less than 800 or missing, 800-950, 950-1000, 1000-1050, 1050-1100, 1100-1150, 1150-1300, or greater than 1300.

or cost of living may also be important, but adding such controls may partially absorb part of the causal impact of colleges—e.g., they impart skills that allow their students to find higher paying jobs in big cities. Controls for more complicated determinants of outcomes like this are also omitted from the analyses presented below.

Estimates of μ_s are the additional earnings students earn by attending school s relative to an omitted "reference" school. Rather than reporting these coefficients, to make the results more easily interpreted, the fixed-effect estimate for each institution is added to the average earnings level for all individuals in the sample. The result is an estimate of the average earnings of students who attended each institution if all institutions enrolled students with average characteristics. Note that the levels of this measure are essentially arbitrary (they shift up or down depending on the types of characteristics used to predict earnings), and it is the difference in these regression-adjusted earnings measures across institutions that captures the estimated difference in quality. Figure 4-3 summarizes how different specifications of the regression model adjust the average earnings of institutions. In the Figure, institutions are grouped into 'bins' in $5,000 increments of the average raw earnings of their students on the x-axis. The markers in the Figure then indicate for each of these groups the average difference between the regression adjusted earnings level (i.e., the predicted annual earnings at an institution if they enrolled students with average characteristics) and the institutions' actual average earnings. This is a measure for how much of the difference in the institutions average earnings is driven by differences in the types of students they enroll relative to the 'average student.' If student characteristics explained none of the differences in outcomes across institutions these dots would all cluster around zero, whereas if only student characteristics mattered the line connecting the dots would have a slope of minus one.

Figure 4-3: Sensitivity of Model to Controls

Predicted minus Actual Earnings

Note: Outliers with earnings above $100,000 not included in figure.
Source: Department of Education, College Scorecard Data

43

The Figure shows that differences in student characteristics are correlated with the earnings outcomes of institutions. [29] In particular, it is clear that institutions with low average earnings differentially enroll students with characteristics that are associated with lower earnings and the opposite is true for institutions with higher average earnings. Looking at the differences across model specifications helps to give a sense for which student characteristics are most important. The first specification including only family income, independent status, and parental education variables explains some of the variation in average earnings outcomes, but does so slightly more at the bottom of the earnings distribution. Adding demographic information in Model 2 results in even more adjustment that is symmetric for high and low-earning institutions. The gender composition of schools, in particular, has an important influence on average student outcomes among this set of variables. The more flexible set of control variables for family income in Model 3 results in a very slight change to the estimates concentrated among schools with high overall earnings levels—presumably owing to the fact that these schools disproportionately enroll students at the highest family income levels. Finally, adding variables meant to proxy for the students' academic background—the number of schools to which the student sent a FAFSA and their average SAT scores—also explains some of the differences in raw earnings for institutions especially at the top of the institutional earnings distribution, where the most selective schools are concentrated.

Figure 4-3 shows that student characteristics influence differences in raw earnings across institutions, and that the differences can be economically meaningful. At the same time, it is important to note that the variation explained by these characteristics is relatively small overall. The blue dots in Figure 4-4 show the relationship between adjusted and unadjusted earnings. While the averaging masks some of the differences that exist for particular institutions, overall the raw averages are very highly correlated (r=0.98) with the adjusted earnings based on the institutional fixed-effect model of causal impact. As noted above, a caution is that if the variables in the specifications presented imperfectly control for factors like students' academic ability (as is almost certainly true), then this method is likely to 'under-adjust' for these differences and thus overstate quality differences across institutions.

To compare the approach above to an aggregate level regression approach, the average earnings for students at each institution (using the same definition described above) are regressed on institution level averages of the control variables used in Model 3.[30] Figure 4-4 shows that relative to estimates based on the individual-level model, the aggregate adjusted earnings exhibit significantly less variation for four-year schools. That is, institutions with low and high raw average earnings are adjusted towards the mean earnings in the sector, suggesting much less of a difference in quality relative to the variation in raw mean earnings. It is important to note that while the adjustment incorporates most available data, it does not control for all important

---

[29] For two-year schools, the model is less sensitive to controls, owing to the fact that there is less variability in the characteristics of students attending two-year schools. As groups of schools or programs become more homogeneous, controls become less important.

[30] There are slight differences in the control variables available for this exercise. In particular, as opposed to separate parental control variables for each parent, the aggregate regression model uses the fraction of individuals where the highest level of education across either parent is middle school, high-school, or college.

information. As an example, aggregate SAT scores might be included to control for differences in academic ability of the student, and geographic controls could be added to control for the impact of cost of living differences, though as explained above it is not obvious such a control is appropriate.

Figure 4-4: Individual vs. Aggregate Adjusted Earnings at 4 Year Schools
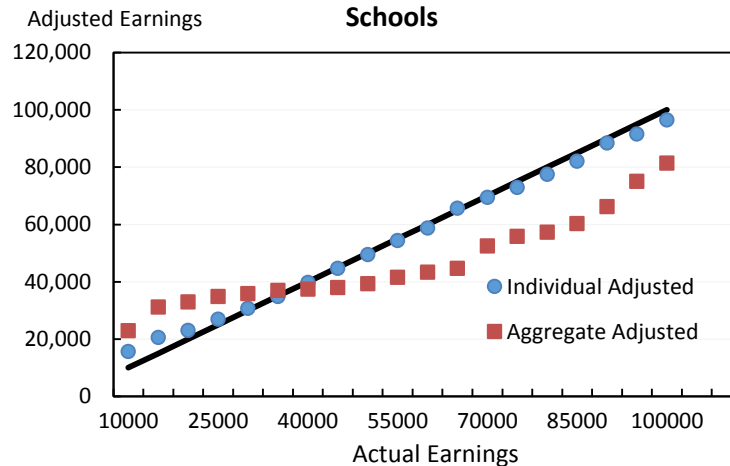
Note: 10 year mean earnings for the 2002 cohort are used.
Source: Department of Education, College Scorecard Data

As discussed above, it is not obvious *ex ante* which method for assessing institutional quality is best. One way to evaluate how reliably these measures predict differences in the causal impact of institutions on student outcomes is to compare the difference in outcomes estimated by each method above to experimental or quasi-experimental estimates of college quality (Kane and Staiger, 2008; Kane et al., 2013; Chetty, Friedman, and Rockoff, 2014). The handful of studies that estimate the impact of college quality based on comparisons of students who barely score above or below SAT test score admissions thresholds might provide a useful starting point for such analyses (Cohodes and Goodman, 2012; Zimmerman, 2014; Goodman, Hurwitz, and Smith, 2015). Recent estimates of college impacts on earnings by Hoxby (2015) may also be a useful benchmark.

V. Data Driven Lessons for Performance Measurement and Accountability

This section uses the data made available by the Administration to discuss some important considerations in measuring the performance of postsecondary institutions and potentially using those measures for accountability. Much of the focus of the discussion uses examples with college earnings outcomes to discuss college quality, as earnings are currently the outcome with the best available data.

Institutional Quality or Selection Effects?

The earnings outcomes experienced by the "typical" student vary widely across different institutions. Figures 5-1a and 5-1b show this variation in two- and four-year schools by depicting the mean earnings for institutions, where institutions are sorted from lowest to highest mean earnings. Each dot in the Figures represents about two percent of all schools in their respective predominant degree level, and indicate the average of the 10 year mean earnings levels at each of these institutions.

Figure 5-1a: Distribution of Mean Earnings at 2 Year Schools
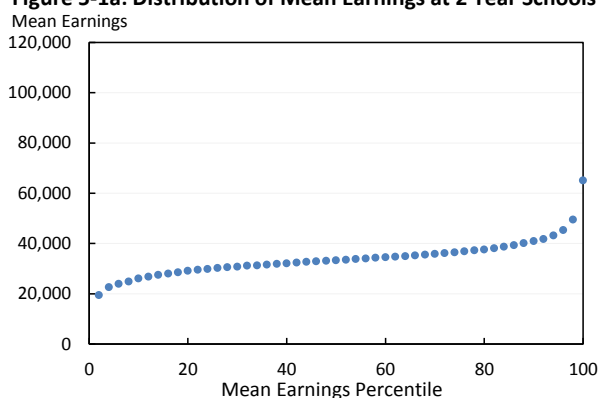
Mean Earnings

Note: 10 year mean earnings for the 2002 cohort are used.
Source: Department of Education, College Scorecard Data

Figure 5-1b: Distribution of Mean Earnings at 4 Year Schools
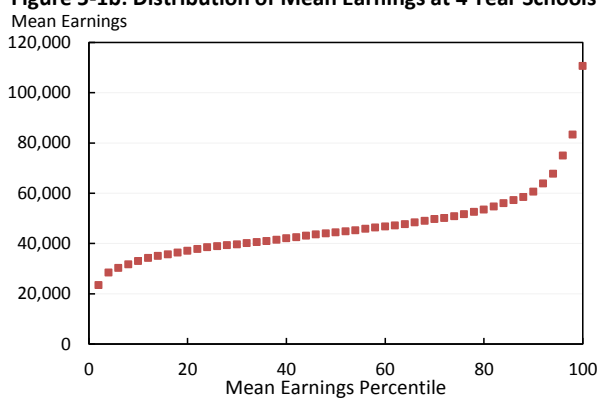
Mean Earnings

Note: 10 year mean earnings for the 2002 cohort are used.
Source: Department of Education, College Scorecard Data

Critics of outcomes-based accountability contend that the variation in outcomes such as these earnings outcomes, or in completion rates, primarily reflect differences in the types of students that colleges enroll, rather than a college's effect on those outcomes. Figures 5-2a and 5-2b show the average fraction of Title IV students in five income categories, among institutions grouped into percentiles of institution-level mean earnings distribution (again for Title IV students), measured 10 years after college entry. Even among Title IV aid recipients there is a dramatic family income gradient across schools with different post-enrollment earnings levels. Among the 10 percent of four-year schools with the lowest earnings, more than two-thirds of students are from families with incomes below $30,000, whereas, in the top 10 percent of four-year institutions, nearly the opposite is true with roughly one-third of low-income students. This gives the impression that factors associated with family income differences may be partially responsible for the differences in student outcomes across these institutions.

Figure 5-2a: Family Income Distribution by Institution Median Earnings Percentiles for 2 Year Colleges
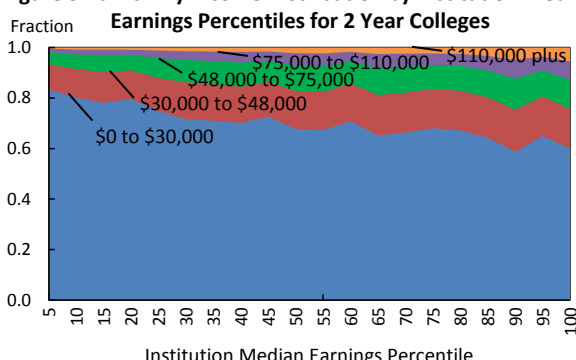
Note: 10 year median earnings for the 2002 cohort and 2012-2013 FAFSA income data are used.
Source: Department of Education, College Scorecard Data

Figure 5-2b: Family Income Distribution by Institution Median Earnings Percentiles for 4 Year Colleges
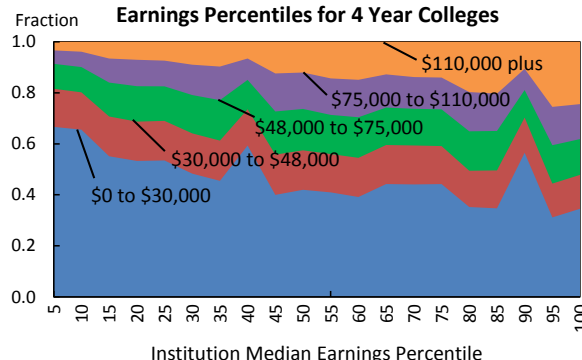
Note: 10 year median earnings for the 2002 cohort and 2012-2013 FAFSA income data are used.
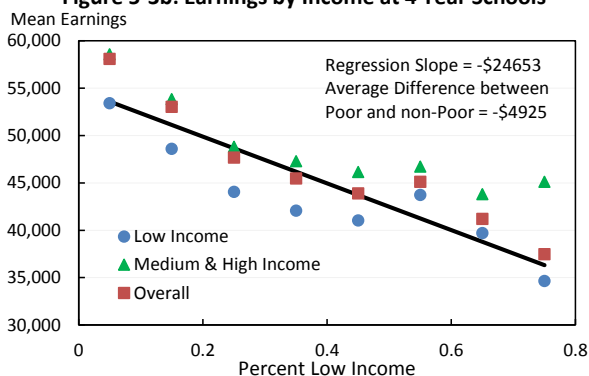Source: Department of Education, College Scorecard Data

Both earnings and completion rates tend to be lower at four-year schools with a higher proportion of students coming from low-income families. However, as shown in Figure 5-3b, average outcomes at these schools tend to be lower for both low-income and high-income students. At two-year schools (Figure 5-3a) a similar pattern is evident, though the difference in the relationship between family income and post-enrollment earnings is slightly more similar within schools and across them than is the case with four-year schools. While tentative, these results, combined with the exploratory estimates of the causal impact of institutions presented in the previous section suggest that differences in institutional quality are important. Research based on more exhaustive data allowing better controls for potential selection effects by Cunha and Miller (2014) and Hoxby (2015) strengthen this conclusion.[31]

Figure 5-3a: Earnings by Income at 2 Year Schools
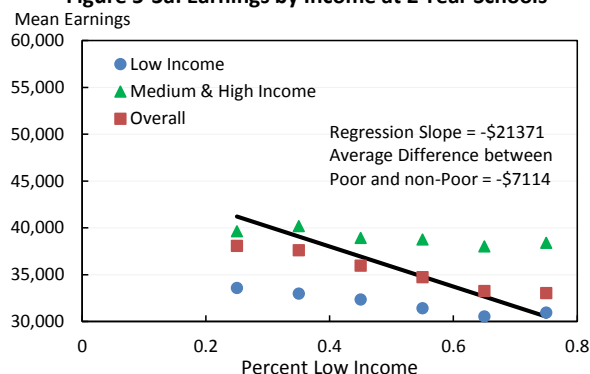
Note: Data for 10 year earnings. Low income defined as less than $30,000, and medium/high income defined as greater than $30,000.
Source: Department of Education College Scorecard Data

Figure 5-3b: Earnings by Income at 4 Year Schools

Note: Data for 10 year earnings. Low income defined as less than $30,000, and medium/high income defined as greater than $30,000.
Source: Department of Education College Scorecard Data,

---

[31] Additionally, as discussed in footnote two, determining when to account for student characteristics in measuring outcomes also needs to be balanced against public policy goals of ensuring that all student populations are served well.

Individual and Social Returns to College Quality

When choosing a school, students must weight all benefits from college (including other dimensions of quality) against the net price of their education. Of course, to properly compare the costs and benefits, the net price (annual cost of attendance less grant aid) needs to be placed on the same time scale as the benefits, otherwise the comparison would underestimate the value of the education.[32] A comparison of net price and earnings, for instance, could show the earning returns the institution generates—and it seems very likely there are some institutions that do generate large earnings returns for students. Additionally, places with large returns may not always charge low prices. Figure 5-4 uses un-adjusted Scorecard data in order simply illustrate this point, showing the relationship between the net price of college for one year and median earnings for four-year institutions, grouped into $5,000 "bins" of median earnings—with between 25 to over 100 institutions in each "bin." For each bin, the Figure shows the average net price of the median institution, as well as the net prices of the institutions at the 10th and 90th percentiles of the net price distribution within the bin. The Figure shows that while institutions with higher earning students tend to have higher net prices, these higher prices may not always offset the earnings advantage.

The schools in the highest earnings category (the far right) on average have earnings levels nearly $60,000 greater than the schools with the lowest average earnings, but the median school's net price is only about $15,000 greater. Because the earnings advantage is enjoyed over at least 30 years and college costs may be paid for just four to five years, earnings benefits likely far exceed the costs (this conclusion would still hold using the most conservative of our exploratory estimates of the difference in causal impact estimates presented in Section IV). A caveat is that schools with higher earnings tend to have higher fractions of their former students go on to attend graduate school. While the earnings gains from subsequent schooling are captured in Figure 5-4, the net price of future schooling—which could be substantial—is not.

---

[32] For illustrative purposes, price and earnings are displayed in simple Scorecard measures. To determine the actual net earnings gain associated with a school, one could subtract the cost of paying for college over 30 years (not including grant aid) from 30 years of earnings after college. Because earnings profiles are not available in this detail, an alternative is to subtract an equivalent one year payment on a 30 year annuity of the total net price of college from one year earnings 10 years after college.

Figure 5-4: Earnings v. Net Price for 4 Year Schools

Net Price
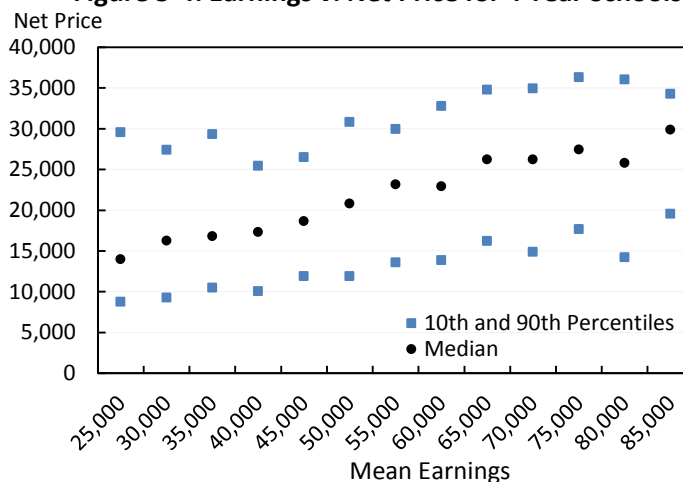
Mean Earnings

Note: Data are from 10 year mean earnings from 2002 and net price from 2013.
Source: Department of Education, College Scorecard Data

Leaving the question of overall net returns aside, Figure 5-4 is suggestive that some schools provide better returns than others. While net price is positively correlated with earnings outcomes (r= 0.33), the Figure shows there is wide variation in the prices students pay to attend colleges with similar earnings outcomes. In most of the bins based on median earnings above, the school at the 90th percentile of the net price distribution (within the bin) has an annual net price that is $15,000 to $20,000 higher than the institution at the 10th percentile of the distribution.

Earnings Outcomes Vary Widely within Institutions, and the Role of Program Mix

While outcomes for the typical student vary substantially *across* schools, there is a great deal of variation *within* schools in the outcomes as well. Students should know that while differences across schools may inform the question of relative school quality, these differences mask a great deal of heterogeneity in the outcomes of students. For example, only about 5 percent of the variation in earnings across students who attend four-year schools is explained by the institution those students attend.[33] This is illustrated in Figure 5-5, which shows the average median and 25th and 75th percentiles of earnings for institutions, where schools are grouped into 50 bins according to their students' median earnings 10 years after entry.

---

[33] This calculation refers to individuals not enrolled in school with positive earnings after the completion time frame. For comparison, among individuals with positive annual earnings aged 25 to 34 and not currently enrolled in school, the number of years of completed education explains about 13 percent of the variation in annual earnings (CEA calculations based on data from the 2014 American Community Survey).
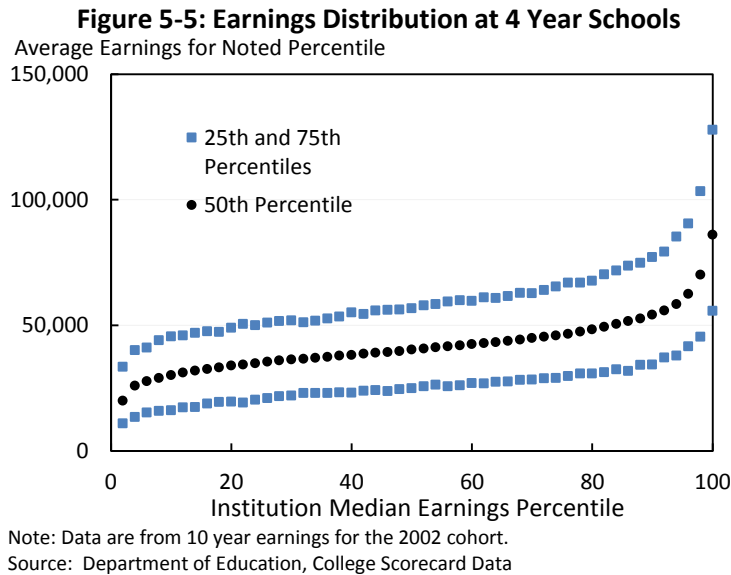
Figure 5-5: Earnings Distribution at 4 Year Schools

Average Earnings for Noted Percentile

Note: Data are from 10 year earnings for the 2002 cohort.
Source: Department of Education, College Scorecard Data

Figure 5-5 shows that a typical student at schools near the 95th percentile of the institution median earnings distribution has earnings that are more than double the average median earnings of schools near the 5th percentile. Still, 25 percent of students attending the latter group of schools (near the bottom of median earnings) earn more than the lowest-earning 25 percent of students at the schools with close to the highest median earnings. The fact that there is so much variation in student outcomes within schools should not be taken as evidence that schools may not matter as much as other factors. The dramatic variation in outcomes within schools does beg the question, however, of whether this variation can be explained by within-school student-level differences in educational experiences that might provide additional useful information to students.

As described in Section II, several studies have documented variation in the earnings outcomes of students within an institution tied to their program of study. While federal data currently do not allow student outcomes to be estimated separately by program of study, several state information systems currently report outcomes separately for students in different programs within the same institution. For example, Figure 5-6 depicts the median year earnings one year post-graduation for Bachelor's Degree recipients in Texas separately by major for each of 35 colleges.[34]

---

[34] The Texas state-level data is based on administrative student records and state unemployment insurance earnings data for graduates only from the website CollegeMeasures.org (College Measures, 2015).

Figure 5-6: Earnings by Institution and Major for Recent Bachelor's Degree Recipients in Texas

Note: Correlation with overall earnings noted in parentheses.
Source: College Measures
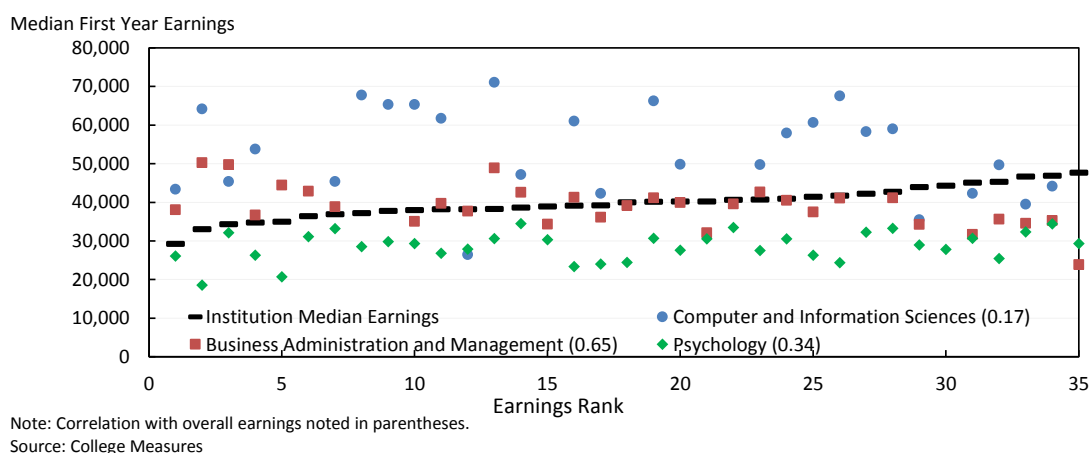
In many cases, the variation in earnings across programs in different institutions mirrors that of the overall institution-level outcome. Naturally, this tends to be the case for large programs at a school (Business Administration and Management is an example in Texas), where the institution-level median earnings appear to be a very good guide to program-level outcomes. For smaller programs, however, the overall institution level median can be misleading about differences in program level outcomes. Figure 5-6 shows that especially for Computer and Information Sciences, the program-specific outcome is quite different from the overall institution outcome. Indeed, 17 of the 27 institutions depicted have higher earnings for students in the Computer and Information Sciences major than students in the same major from the institution with the highest overall median earnings (a school with a heavy focus on health fields). This highlights that for students committed to studying in a particular field, program-specific outcomes provide important additional information that may be obscured by institution-level data.
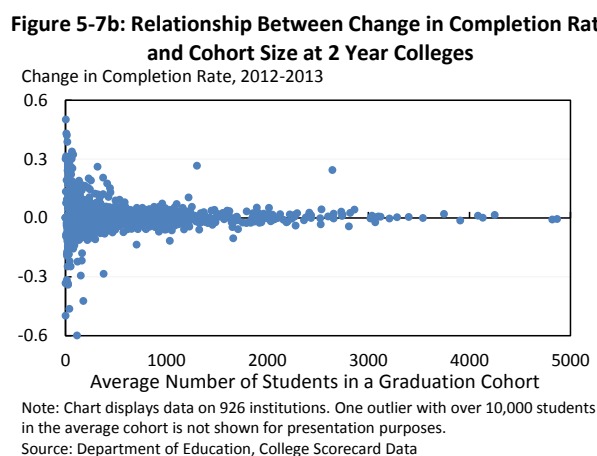
Institution-level data are still likely to be helpful, however, as many students may enroll without a clear sense for what major they will pursue or switch majors after enrolled. Moreover, for accountability and benchmarking purposes, summary measures of institutional performance may be necessary. A challenge with developing program-level data, on the other hand, is how to characterize the field of study of students that do not complete a degree in that field. Moreover, controlling for different selection criteria into particular majors within institutions pose additional challenges to estimating the causal impact of programs on outcomes. At this time, the Department can only report program-level outcomes using data collected under the Gainful Employment regulations. However, beginning in 2012 the Department began collecting program-level information for all Title IV recipients in NSLDS and will be working to make that information available to the public.

Reliability of Performance Measures and Information on Smaller Schools

An important consideration in evaluating performance measures is whether they reflect stable aspects of institutional quality, or whether they vary substantially from year to year. Performance measures may be variable in smaller institutions, where a handful of students' outcomes may

have an outsized influence on the average institution outcome. Measures might also fluctuate if there are idiosyncratic factors that influence outcomes in a particular year, such as the closing of a large local employer, a natural disaster, or a short-lived but effective program to boost completion rates (Kane and Staiger, 2002).

Figures 5-7a and 5-7b illustrate this phenomenon using reported graduation rates in IPEDS. As can be seen in the Figures, the variation in completion rates is much higher among institutions with relatively small numbers of students in the graduation cohort. Figure 5-7b also shows that smaller institutions are much more likely to have large swings in performance from year to year. This raises the concern that performance metrics for smaller schools may be unreliable, but also that institutions with the highest and lowest scores on a given metric are likely to be smaller institutions whose performance measures tend to vary most from year to year.

Figure 5-7a: Relationship Between Completion Rate and Cohort Size at 2 Year Colleges

Note: Chart displays data on 970 institutions.
Source: Department of Education, College Scorecard Data

Figure 5-7b: Relationship Between Change in Completion Rate and Cohort Size at 2 Year Colleges

Note: Chart displays data on 926 institutions. One outlier with over 10,000 students in the average cohort is not shown for presentation purposes.
Source: Department of Education, College Scorecard Data

To mitigate volatility in performance measures due to both small cohort sizes and idiosyncratic influences on outcomes, the new data made available on the College Scorecard are based on multiple cohorts of students. For earnings data, measures are based on two successive entry cohorts, and statistics with fewer than 50 students are suppressed. For IPEDS and NSLDS completion rates, repayment rates, and cumulative loan debt, two successive cohorts of students are combined.[35] For consumer measures, among institutions where there were fewer than 30 students in the combined cohorts, the measure was created based on a four-year cohort instead.[36] The result, discussed below, is that the reported measures are more reliable than would be the case for single cohort measures. A cost is that this masks changes in performance over time, but given that most measures are observed with a lag of at least 5 or 6 years after students first enter an institution, the incremental cost of averaging over cohorts seems low.

---

[35] For example, the most recently available 6-year completion measure in NSLDS is based on students who first entered an institution in either the 2007 or 2008 award years.

[36] In the four cohort case, 2005 and 2006 were also added. The roll-up of two successive two-year cohorts of data was only done for Department of Education data based on IPEDS or the NSLDS. Data on earnings are presented in two-year cohorts only, and suppressed when there are fewer than 50 individuals with valid earnings in a cohort. For all measures, further disclosure protections are sometimes applied, if for example subgroup cell sizes are small.

Data that has only been suppressed for privacy reasons is available to researchers and other analysts, however, and these users are advised to be wary of statistics based on small samples. There is substantial correlation between outcome measures over time. Table 5-2 below shows the correlations between key performance measures for successive cohorts of IPEDS graduation data, and successive two-year cohorts for earnings and repayment data. For example, columns marked "one" show the correlation between successive two-year cohorts—e.g., the 1997 and 1998 entry cohort's median earnings eight years post-college entry, with the 1999 and 2000 entry cohorts' median earnings. The column marked "two" denotes the correlation of the 1997 and 1998 entry cohort metric with the 2001 and 2002 entry cohorts' metric. It also characterizes the stability of the measures in another way, by showing the fraction of institutions identified as the top or bottom 10 percent of performers that remain in the top or bottom cohort in subsequent cohorts.

Overall, outcome measures tend to be stable over time for both two- and four-year schools, with many of the correlations across successive cohorts above 0.9. Even comparisons two cohorts apart remain very highly correlated. The Table also shows the fraction of schools that remain in the top or bottom 10 percent of performers. Stability in this measure reflects both the correlations reported in the first two columns of data in Table 5-2, and the overall variation in the measure. Even where the year-to-year correlation of the measure is very high, if many institutions have similar outcomes near the top or bottom of the distribution then small fluctuations might move many institutions above or below the 10 percent threshold. For earnings and completion measures, between about 25 to 30 percent of institutions identified in the top or bottom 10 percent at the first measure are not in that category two measurements later. Results are more stable at the top of the distribution than the bottom, and for four-year schools. The composition of institutions with repayment rates in the top and bottom 10 percent tend to change the most among the metrics examined, despite relatively stable correlations across the distribution. As shown below, this reflects the fact that there is less variation in these measures among schools with high repayment rates.

Table 5-2: Correlation Between Years of Outcome Measures						
Measure	Correlation		Fraction Remaining in Top 10%		Fraction Remaining in Bottom 10%	
Cohorts Distance	One	Two	One	Two	One	Two
IPEDS Completion (2 Year)	0.90	0.89	0.73	0.71	0.70	0.74
IPEDS Completion (4 Year)	0.88	0.88	0.85	0.86	0.75	0.74
10 Year Earnings (2 Year)	0.95	0.93	0.78	0.74	0.81	0.76
10 Year Earnings (4 Year)	0.97	0.94	0.89	0.92	0.83	0.78
3 Year Repayment (2 Year)	0.74	0.67	0.82	0.72	0.63	0.60
3 Year Repayment (4 Year)	0.85	0.82	0.67	0.60	0.67	0.52
Note: Correlations are distanced one and two cohorts apart using the most recent years of data available, where a cohort represents a year or rolling average completely distinct from the one on either side of it. Rolling averages are used for all measures except for completion. IPEDS completion rates correlate 2011 data with 2012 and 2013 data. Earnings correlate the rolling average of 1997-1998 data with 1999-2000 and 2001-2002 data. Repayment correlates the rolling average of 2006-2007 data with 2008-2009 and 2010-2011 data. Correlations and fractions are calculated separately for two and four year schools.						
Source: Department of Education, College Scorecard Data						

While the cohort to cohort correlations in Table 5-2 suggest a high degree of reliability for most performance metrics overall, users of the data should still be cautious about using data with relatively few individuals included in any given metric. Figure 5-8 illustrates the concern using three-year repayment rates for four-year schools, reporting the correlation across combined cohorts. The Figure shows the correlation across "rolling" two-year cohorts with the repayment rate for students entering repayment in 2006 or 2007 broken down by cohort size. The points above "1" on the x-axis depict the correlations of these base year repayment rates with the repayment rates for students entering repayment in 2007 and 2008. The Figure shows that the reliability of the repayment rate measure based on small cohorts can be fairly low—falling below 0.2 over a four year span.

Figure 5-8: Correlation in 3 Year Repayment Rates Over Time at 4 Year Colleges

Note: Data for 2008-2011 cohorts compared to the 2007 cohort, using rolling averages. Student size corresponds to cohort size.
Source: Department of Education, CollegeScorecard Data
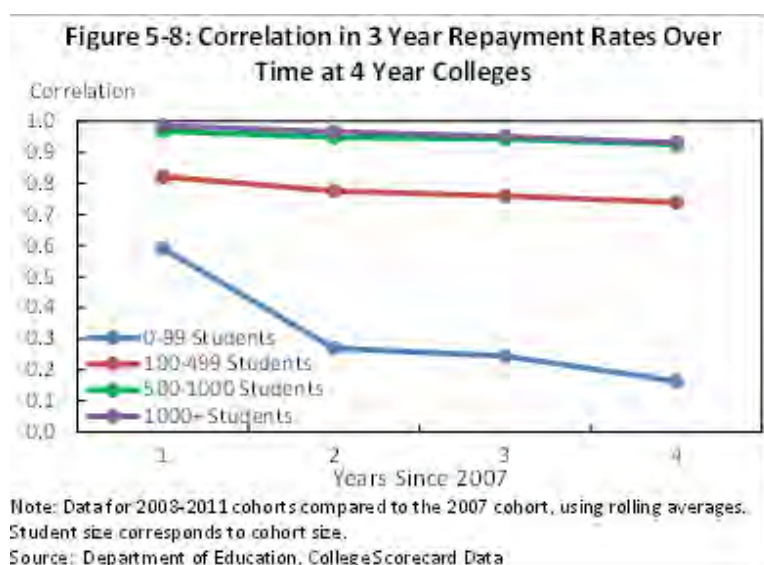
For measures based on student subgroups, or when considering program-level or other data that would involve smaller sets of students, reliability may be an even bigger concern. To improve reliability, analysts might consider combining groups or cohorts to yield bigger sets of students, using several successive measures for an outcomes, or using more sophisticated techniques.[37]

How do Various Dimensions of Performance Relate?

The key measures of student success available in the data are graduation rates, post-enrollment earnings, and loan repayment. These measures may capture different dimensions of college performance and may proxy for different goals students have for their higher education. Completing a degree may be seen as important per se, or proxy for the accumulation of specialized knowledge, regardless of how much a student goes on to earn after graduation. Similarly, loan repayment rates may capture variation in the affordability of a college, or perhaps

---

[37] For example, McClellan and Staiger (2000) propose a method of "filtering" estimates, essentially using the history of each of several measures to optimally predict each outcome measure accounting for noise due to sampling error.

the degree to which it assists students with financial literacy to be sure they're able to manage their debts, even among institutions where students experience similar earnings outcomes.

At the same time, these measures undoubtedly aim to capture similar dimensions of institutional quality as well. If institutions have a large causal impact on earnings, we would expect graduation rates to be strongly associated with earnings. And since having higher earnings contributes to students' ability to repay their loans, we would expect both completion and earnings to be related to repayment rates.

Figures 5-9a-c depict the relationship among these three performance measures separately across two- and four-year schools. Figure 5-9a shows the relationship between completion and earnings. For four-year schools, there is a positive correlation between earnings and completion overall, but the relationship is non-linear. There is essentially no correlation between earnings and completion among schools with a completion rate below 50 percent (about 46 percent of four-year schools), but then a strong positive relationship among schools above that mark. Among two-year schools, the correlation between completion rates and earnings is very low (.07). This lack of correlation at both two-year schools and four-year schools with low completion rates raises questions about the value of completion rates as an indicator of quality. Of course, as noted already, it is possible that completion rates capture college quality in dimensions that are uncorrelated with the institutions' earnings (note the Figure would look similar if we used regression adjusted earnings).

Figure 5-9a: Relationship Between Earnings and Completion Rate
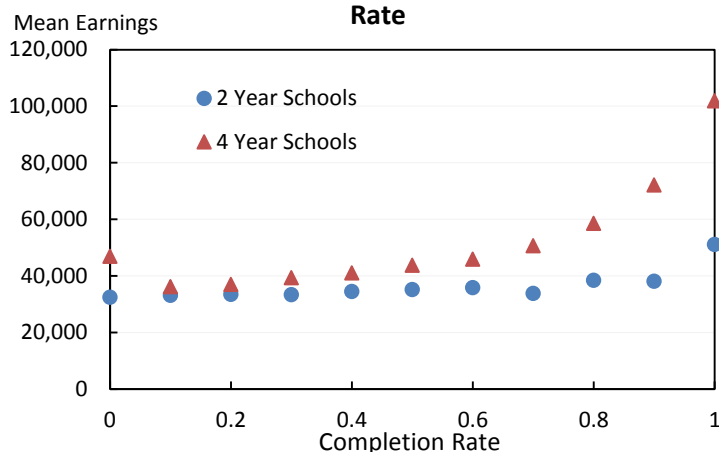
Note: Chart uses 10 year mean earnings and the 150% completion rate from IPEDS.
Source: Department of Education, College Scorecard Data

Figure 5-9b provides similar evidence showing the relationship between repayment and completion rates. While there is a positive correlation between repayment and completion at four-year schools, the relationship is very weak for schools with lower and higher completion rates. That is, higher completion seems associated with higher repayment rates mainly for institutions with completion rates between about 20 and 60 percent. For two-year schools, completion rates appear largely unrelated to repayment rates, calling into question what types

of quality information might be reflected in completion rates. These results should not be taken as endorsement of the conclusion that completion has no causal impact on earnings or repayment outcomes. More careful micro-econometric studies have found evidence of earnings gains related to credit accumulation (Jacobson, LaLonde, and Sullivan 2005), and positive earnings impact of completing two- and four-year degrees conditional on enrolling (Kane and Rouse (1995); Jaeger and Page (1996)). Further analyses will be required to unpack this relationship.

Figure 5-9b: Relationship Between Repayment and Completion Rate

Average Repayment Rate
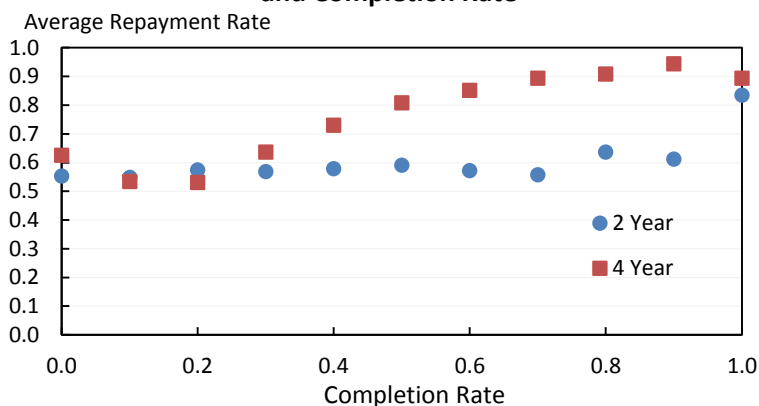
Completion Rate

Note: Repayment data are for the 2010 and 2011 exit cohorts using 3 year repayment rates. Completion data are for 2013 using 150% rates from IPEDS.
Source: Department of Education, College Scorecard Data

Finally, Figure 5-9c shows the relationship between institutions' repayment rates and the median earnings of students 10 years after entry. Repayment and earnings are strongly related at both two- and four-year schools. Again, however, the correlation between the two is stronger at schools with lower median earnings: Above about $50,000, the correlation becomes rather weak. Repayment rates are uniformly high for these top 10 percent of schools, while earnings tend to vary more greatly.

**Figure 5-9c: Relationship Between Repayment Rate
and Median Earnings**

Average Repayment Rate
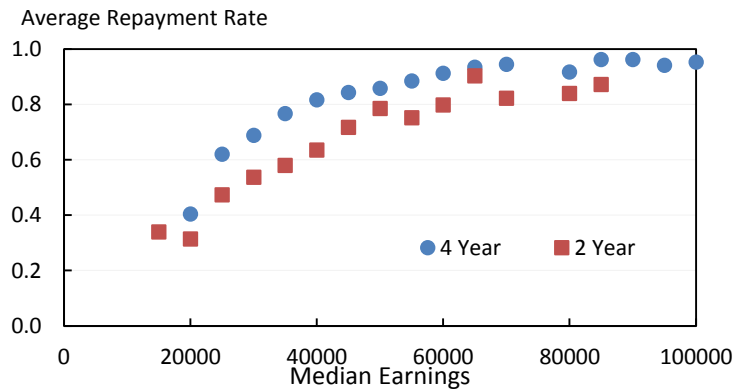
Note: Chart compares the 3 year repayment rates for the 2010-2011 pooled cohort
to the 10 year earnings for the 2002 cohort. Outliers with earnings greater than
$100,000 excluded.
Source: Department of Education, College Scorecard Data

Tables 5-3a and 5-3b provide the correlations across a broader set of measures for two- and four-year schools, respectively. The results in the Tables reinforce the broad conclusions presented above. For two-year schools, the correlations across any of the completion measures—including those that incorporate transfer rates from NSLDS—to either repayment or earnings outcomes are quite low. For four-year schools, there are higher correlations across measures.

| Table 5-3a: Correlations Between Key Outcome Measures at 2 Year Schools								
	Regression Adjusted Mean 10 Year Earnings	Mean 10 Year Earnings	Mean 6 Year Earnings	% Earnings >25K after 6 Years	3 Year Repayment Rates	Regression Adjusted NSLDS Completion Rates (+ Transfer)	NSLDS Completion Rates (+ Transfer)	IPEDS Completion Rates
Regression Adjusted Mean 10 Year Earnings	1.00							
Mean 10 Year Earnings	0.94	1.00						
Mean 6 Year Earnings	0.82	0.88	1.00					
% Earnings >25K after 6 Years	0.69	0.80	0.92	1.00				
3 Year Repayment Rates	0.33	0.41	0.47	0.53	1.00			
Regression Adjusted NSLDS Completion Rates (+ Transfer)	0.03	0.12	0.17	0.16	0.14	1.00		
NSLDS Completion Rates (+ Transfer)	-0.04	0.08	0.22	0.24	0.23	0.81	1.00	
IPEDS Completion Rates	-0.02	0.07	0.13	0.12	0.10	0.75	0.79	1.00
Note: Data are for the most recent cohort of data available for the noted variable.								
Source: Department of Education, College Scorecard Data								

Table 5-3b: Correlations Between Key Outcome Measures at 4 Year Schools								
	Regression Adjusted Mean 10 Year Earnings	Mean 10 Year Earnings	Mean 6 Year Earnings	% Earnings >25K after 6 Years	3 Year Repayment Rates	Regression Adjusted NSLDS Completion Rates	NSLDS Completion Rates	IPEDS Completion Rates
Regression Adjusted Mean 10 Year Earnings	1.00							
Mean 10 Year Earnings	0.97	1.00						
Mean 6 Year Earnings	0.90	0.92	1.00					
% Earnings >25K after 6 Years	0.74	0.77	0.83	1.00				
3 Year Repayment Rates	0.47	0.51	0.46	0.60	1.00			
Regression Adjusted NSLDS Completion Rates	0.45	0.47	0.39	0.36	0.38	1.00		
NSLDS Completion Rates	0.55	0.58	0.49	0.51	0.61	0.73	1.00	
IPEDS Completion Rates	0.59	0.62	0.50	0.51	0.71	0.58	0.76	1.00
Note: Data are for the most recent cohort of data available for the noted variable.								
Source: Department of Education, College Scorecard Data								

Short-Term or Long-Term Outcomes?

As discussed in section I, there is an inherent tradeoff in choosing when to measure outcomes. The choice of time-horizon depends on what aspect of institutional performance one hopes to measure. For rating institutional quality to inform college choice, longer-run measures of earnings may better measure lifetime earnings, while short-term earnings provide information about whether a student is able to get any job after college or can successfully repay his loans soon after completing. Chetty et al. (2014) show the correlation of annual earnings and lifetime earnings stabilizes by about age 30, motivating the choice to focus on earnings measures 10 years after entry since individuals in the earnings cohorts will nearly all have reached the age of 28 by then. They also better convey the long-run earnings *levels* students can expect if they attend various institutions, providing a better sense of the benefits relative to not attending school. Figure 5-10 shows that earnings rise in real terms for students in their early career, increasing by about $5,000 on average between their 6th and 10th year after entering college for two-year public school students, and about $8,000 on average for four-year public school students.

Figure 5-10: Mean Earnings over Time

Mean Earnings
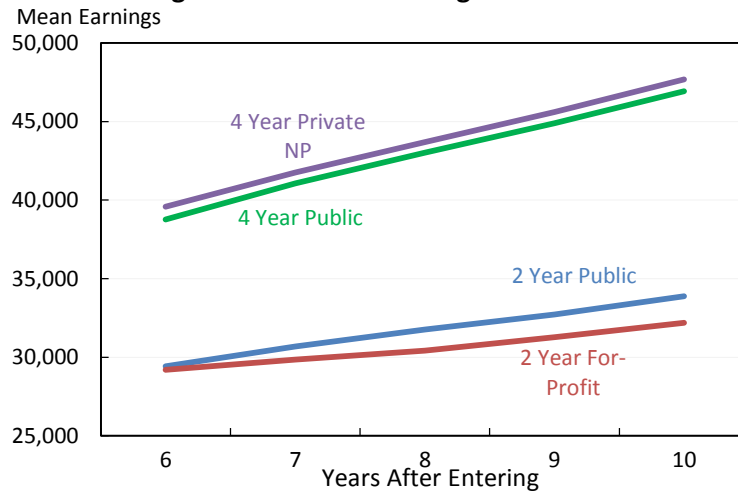
Note: Earnings are listed in real 2014 dollars. Simple, unweighted averages are used.
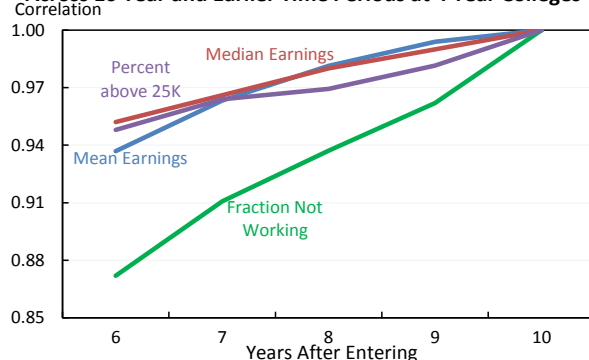Source: Department of Education, College Scorecard Data

Another way to think about this is to examine correlations in outcomes measured at different points in time. In the data, measures of labor market success are estimated in each year from six to 10 years after students first enroll at an institution. Figures 5-11a and 5-11b show that the correlation across these measures is very high. For both two- and four-year institutions, the correlation of mean and median earnings, and the fraction of former students earning above $25,000 per year across metrics based on different time horizons, are all above about 0.94. The correlation in the fraction of students not currently enrolled who are not working is only slightly lower, but still above 0.87. Thus, for conveying the relative performance of the institution (and not necessarily the levels), little seems to be lost by using more recent data. For use in accountability schemes, shorter-run data may be more appropriate. If institutional offerings and the relevant student outcomes may be changing rapidly, then data with as little lag in measurement as possible may provide the most accurate assessment of current performance.

Figure 5-11a: Correlations Between Key Earnings Measures Across 10 Year and Earlier Time Periods at 2 Year Colleges
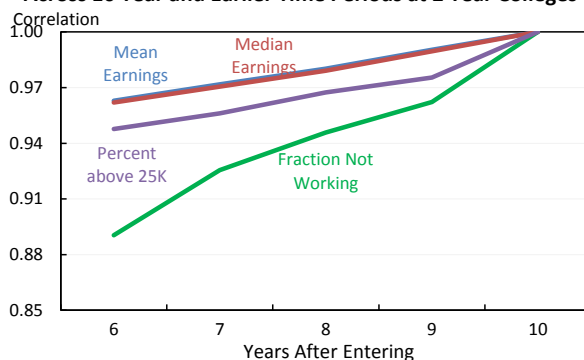
Correlation

Note: Simple, unweighted averages are used in the collapsed data.
Source: Department of Education, College Scorecard Data

Figure 5-11b: Correlations Between Key Earnings Measures Across 10 Year and Earlier Time Periods at 4 Year Colleges

Correlation

Note: Simple, unweighted averages are used in the collapsed data.
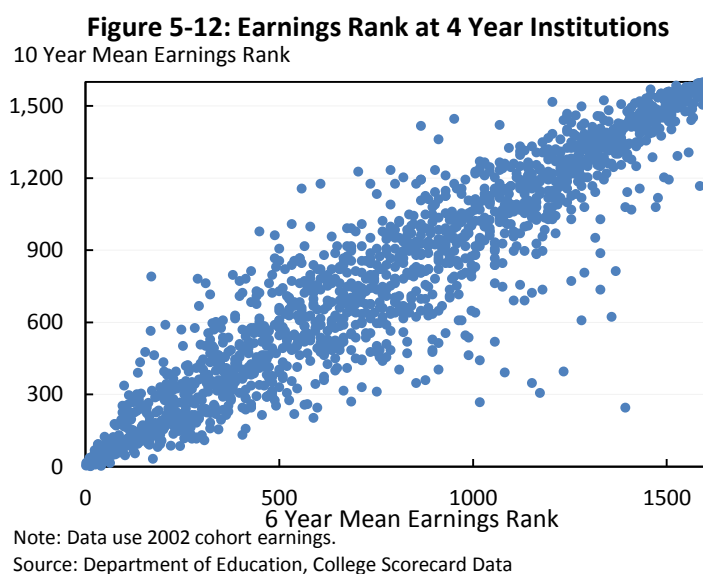Source: Department of Education, College Scorecard Data

While a measure may be highly correlated across different points of time, rank order may still be quite sensitive to the choice. If measures are used to order institutions, shifting between

measures can have a large impact on rank order. Since many institutions are clustered in a narrow band of median earnings, shifting between earnings measures can result in sometimes large changes in the rank of an institution relative to others. These changes in rank tend to be more dramatic for institutions near the middle of the institutional earnings distribution, since the differences in earnings among many institutions there tend to be small (see Figures 5-1a and 5-1b). Figure 5-12 illustrates this, showing how institution ranks using median earnings change based on whether the measure six or 10 years post-enrollment is used.

Figure 5-12: Earnings Rank at 4 Year Institutions

10 Year Mean Earnings Rank

6 Year Mean Earnings Rank

Note: Data use 2002 cohort earnings.
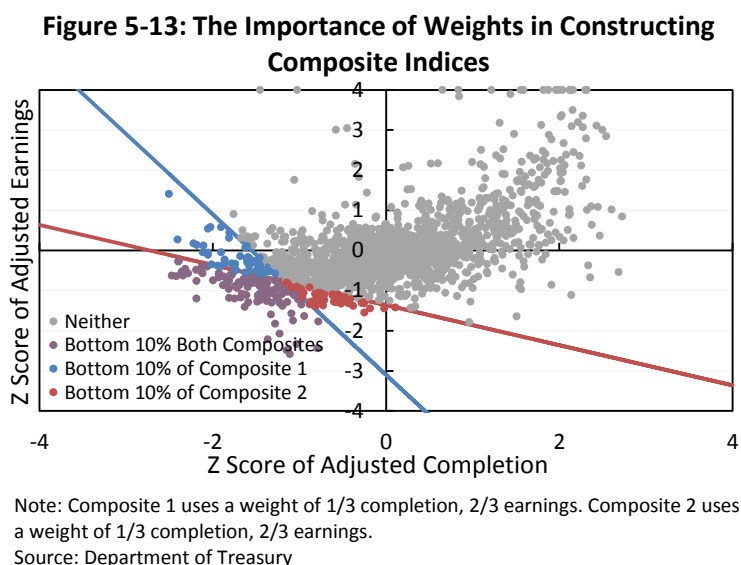Source: Department of Education, College Scorecard Data

The Challenge of Creating Summative Measures of Performance

Separate metrics of institutional performance along each measured dimension of access, affordability, and student outcomes provide potential students and their families with a greater amount of information to form their own assessments of which college or university is the best choice given their goals. On the other hand, reviewing each metric along a broad range of information separately puts the burden of synthesizing that information on potential students, who may have difficulty making these tradeoffs without proper support. To address this, it might be desirable to create measures that combine several indicators into an overall quality rating, or to design a small number of measures that capture performance across a few domains (e.g., access or student outcomes). As the Tennessee Performance Based Funding example in section II illustrates, combining performance across a number of indicators is a necessary step in most consequential accountability schemes.

Measures can be combined in a variety of ways, depending on the particular context. In some cases a natural way of combining different data elements may present itself. For example, it might be useful to help students compare earnings and the net cost of attendance across colleges presenting a single estimate of the net return to attending each college. To compute the net return, we would ideally combine earnings and net price data to estimate the present value of the additional earnings a student would gain by attending an institution relative to the present

value of the net price paid. This is conceptually straightforward, though empirically challenging since we don't observe the full earnings history for students (or sometimes future education and net prices paid).

More frequently, different dimensions of performance—measured in varying units—need to be combined to create a composite index; the best procedure for doing so is often not obvious. For example, an index of student outcomes could be formed based on the earnings and completion of students at an institution. Similarly, policy-makers might value an index that simultaneously combines measures of access, affordability, and outcomes. One way to construct such indices is to transform each variable into "z-scores" (deviations from the mean outcome in their sector, relative to the standard deviation of those outcomes), and to form a weighted average of the separate scores. The resulting index can be a useful way to summarize institutional performance across a variety of dimensions.

Figure 5-13: The Importance of Weights in Constructing Composite Indices

Note: Composite 1 uses a weight of 1/3 completion, 2/3 earnings. Composite 2 uses a weight of 1/3 completion, 2/3 earnings.
Source: Department of Treasury

However, an important challenge with this approach is that there is frequently no clear optimal way to weight each performance indicator. This matters, since an institution's relative performance can be sensitive to the weights chosen. To illustrate, Figure 5-13 depicts the z-scores for adjusted earnings and graduation rates for 1,600 four-year institutions. The lines in the picture delineate the set of institutions identified in the bottom 10 percent (i.e., institutions to the left of the lines) of performance in student outcomes based on two separate composite indices: One that assigns 1/3 weight to completion and 2/3 weight to earnings, and another which reverses those priorities. The index weighted more heavily towards completion (Composite 1 in the Figure) identifies the blue and purple dots as low performers. The Figure illustrates, however, that as the weights shift to emphasize earnings over completion, the set of institutions identified as low performers changes. In this case, 47 of the 160 schools are identified as low performers by Composite 1, but are not categorized as low performers by Composite 2. Since classifications reflect the weights assigned, future work is needed to understand how the relative importance

of each measure should be determined—an issue that will depend on the intended audience and impact of the measures.

Conclusion

The choice of whether and where to attend college is one of the most consequential decisions that most Americans will make. Decades of research have shown large returns to higher education in terms of labor market earnings, health, and happiness. Moreover, the amount students must pay to attend college, the likelihood they will complete their degree, their success in the labor market, and the likelihood they will pay off their loans all depend importantly on the institution where they choose to pursue their education.

These large differences in outcomes across institutions make it clear that students and their families need comparable and easily accessible information about institutions to help them make the best choices to further their goals. While no system can hope to cover the full range of outcomes important to every student, all students should have access to basic information about the financial consequences of their decisions so they can pursue their passions with a clear understanding of the economic tradeoffs.

The new College Scorecard provides just this. It introduces the first nationally comprehensive data on students' post-enrollment earnings, measured for a consistently defined set of students at nearly all post-secondary institutions in the United States. The data also include new measures of the cumulative debt levels of borrowers who complete their degree at each institution, giving prospective students the best information available on the total amount of debt they might accrue and how much they can expect to pay from their future earnings to service their loans. And, to give students a sense of whether typical borrowers at each institution are able to manage their debt, an improved measure to capture students' success in repaying their loans has also been created. Finally, with the help of the higher education community to improve existing reporting requirements, a newly developed measure of completion has the promise to better capture how well institutions promote the degree attainment goals of all of their students.

At the same time, the data provide new tools for higher education officials and policymakers concerned with how well institutions are providing access to diverse groups of students. In addition to highlighting the fraction of lower-income students that institutions enroll, the data also reveal average student outcomes for several student subgroups, including those defined by family income, first-generation status, gender, and age. With this new data, stakeholders can better assess the extent to which institutions are promoting the success of all their students and target support where necessary.

The new College Scorecard data will help students to better select the institutions that best further their goals and will help institutions to benchmark and improve their performance. Still, providing information on its own may be insufficient to alter institutional performance everywhere. Many states and the federal government are pursuing policies to introduce accountability mechanisms for postsecondary institutions. These systems aim to ensure students get value from their education, are not enrolled under deceptive pretenses, or are not left with debt that is unmanageable. Thoughtful consideration of the incentives created by performance metrics and their potential for unintended consequences is crucial, the data described in this

report offer new tools to design systems that simultaneously promote access, affordability, and student success.

Accurate and reliable data are crucial to support better consumer choice, institutional benchmarking, and accountability. While no data are perfect, the newly available data represent a large step forward over other publicly available information in painting a more comprehensive picture of how well postsecondary institutions are providing value to the students they serve. By making federal data on the performance of U.S. institutions widely available to the public, to academics, and to the institutions themselves, the Administration hopes the increased transparency that these data afford will stimulate further progress in measuring both student success and the causal impact of colleges.

Technical Appendix: Overview of the Measures Used

This appendix presents considerations for the outcome and cost measures of the College Scorecard data introduced in Section III of the "Using Federal Data to Measure and Improve the Performance of U.S. Institutions of Higher Education" report. While an overview of construction and cohort definitions is provided in the main report, this appendix provides more detail about the characteristics of the data and description of various metrics, and some factors data users might review prior to using the data.

Five areas of college performance measures are discussed below: earnings, completion, cost, debt and repayment and access. Each content area section explores different topics related to measurement in greater detail than in the main report. First, the earnings section provides summary statistics about the distribution of earnings for college entrants using the Scorecard earnings data and outlines the benefits of using Scorecard data rather than Unemployment Insurance (UI) data from states to measure earnings. Second, the completion section discusses the National Student Loan Data System (NSLDS) data as a source for completion outcomes and compares this data to other data sources. Third, a section on costs describes the components of college cost, with attention paid to the differences between various cost measures. Fourth, the debt and repayment section presents details about the NSLDS data used to track student borrowing and provides a comparison of default rates and repayment rates. Finally, the access section presents and compares a number of metrics that may be constructed with existing data to measure accessibility for disadvantaged students.
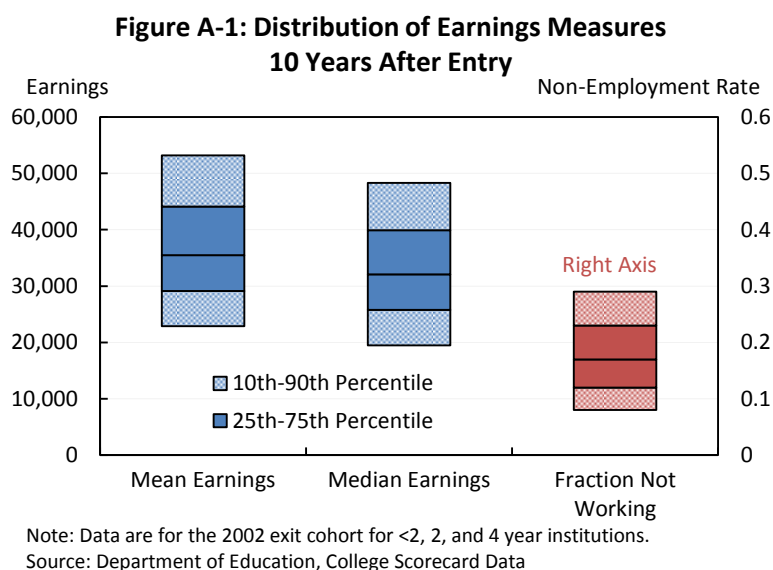
The new College Scorecard data can empower consumers and policymakers involved in the college decision making process. In particular, the data are well suited to bring transparency to the outcomes and costs faced by disadvantaged students in colleges across the country. However, it is essential for users to understand the strengths and limitations of the outcome and cost metrics available in order to make beneficial and well-informed decisions.

Earnings

While post-college earnings represent only one dimension of labor market success, they are an important indicator for students and a signal of institution quality. Post-enrollment earnings are one of few meaningful indicators available for nearly all institutions that are reported on a common scale. Even though students enroll in diverse programs of study, their earnings reflect the labor market's valuation of the human capital acquired in school.

The simple measures outlined in Section III of the main report—of mean earnings, median earnings and fraction not working—provide an overall assessment of the degree to which past attendees of an institution are able to find good jobs that pay well. These measures are calculated using administrative IRS tax records, a resource that has several benefits when compared to other earnings data sources, including state UI records.

Using data for the 2002 entry cohort, Figure A-1 shows that the mean earnings 10 years after entry across all colleges range from $29,100 at the 25th percentile to $44,100 at the 75th percentile, with a median of $35,500 (all dollar amounts are presented in current 2014 dollars).[38] Median earnings are less influenced by individuals with very high earnings; they are thus lower, with an interquartile range from $25,800 to $39,900 and a median of $32,100. The fraction of students not enrolled and not working has an interquartile range from 12 percent to 23 percent with a median at 17 percent.

Figure A-1: Distribution of Earnings Measures
10 Years After Entry

Note: Data are for the 2002 exit cohort for <2, 2, and 4 year institutions.
Source: Department of Education, College Scorecard Data

It is important to note how well IRS data compare to administrative state unemployment insurance (UI) data on quarterly earnings. The state UI data have some advantages and disadvantages relative to the IRS data. An advantage—though not related to the underlying data—is that state UI based earnings are reported for all students rather than Title IV students only. On the other hand, state UI data are limited in that only students who work in the same state as the institution can generally be matched (though efforts to match UI records across states are underway), and certain kinds of earnings (e.g., self-employment earnings) and occupations not covered by the UI system are excluded. To assess the degree to which the latter two limitations might result in biased earnings measures, estimates of mean and median earnings were created imposing the limitations of the UI on the earnings data used for the scorecard. In other words, institutional mean and median earnings were re-estimated for the same cohorts of students but restricting data to individuals living in the same state where they attended college and omitting self-employment earnings.[39]

The comparison of the restricted state UI data and IRS earnings data has several interesting features. First, limiting the data to students who are living in the same state can ignore the outcomes of a very high fraction of students due to migration. On average across institutions the

---

[38] We use 10 year measures for illustrative purposes.
[39] Unemployment insurance data are also restricted to Title IV students for comparison purposes, unlike true unemployment insurance data.

data constructed with UI restrictions cover only about 62 percent of students in scorecard data. This raises the concern that these data could be biased, depending on the nature of selection into working in the state versus elsewhere for each college. Despite this, however, Figure A-2a shows that overall the difference in average earnings measures is fairly low: the average log difference between the two measures is less than one percent The errors are largely symmetric, with about 14 and 10 percent of institutions having measured UI earnings that are more than five percent below and above the scorecard estimates, respectively. Figure A-2b shows that relative to the Scorecard analysis using IRS earnings data, UI data result in lower earnings estimates when a lower fraction of former students are covered. This suggests that on average, individuals moving out of state tend to have higher earnings than those remaining in state. It is beyond the scope of this report to assess whether this reflects higher mobility across states for higher-income workers, a concentration of colleges relative to employers of college-educated workers in higher-wage states, or some other factor.

UI data based earnings estimates also appear to understate average earnings relative to IRS earnings at schools with higher earnings measured using IRS. For institutions in the bottom 60 percent of institutional mean earnings, the result is that UI data appear slightly biased upwards on average by a few percent. At institutions with average earnings near the top of the distribution, the result is the opposite—institutions in the top decile have on average 5 percent lower UI data-based earnings. This may result from the restriction of the data to in-state earnings and the increased mobility of high-skilled workers, shown in Figure A-2c. Overall, there are large gaps in the number of students covered by UI data, but despite these differences, estimates of average and median earnings based on UI and national tax data are quite similar. Still, for a small fraction of institutions, quite sizeable differences exist. Further work is necessary to characterize the sources of these differences, but the fraction of students who work out-of-state and the level of earnings overall drive some of the bias in UI data-based estimates.

Figure A-2a: Institutional Level Difference of UI Earnings Minus Tax Earnings by Earnings Difference (UI minus Tax)
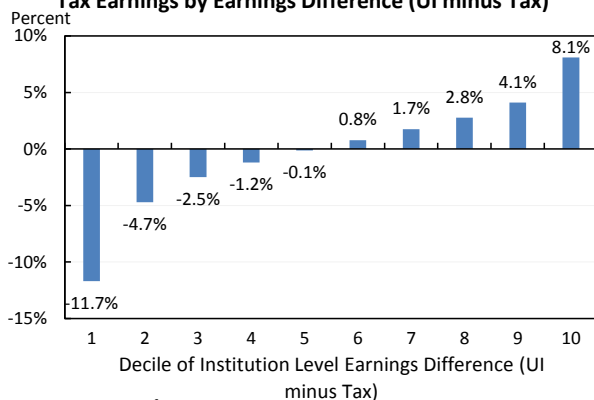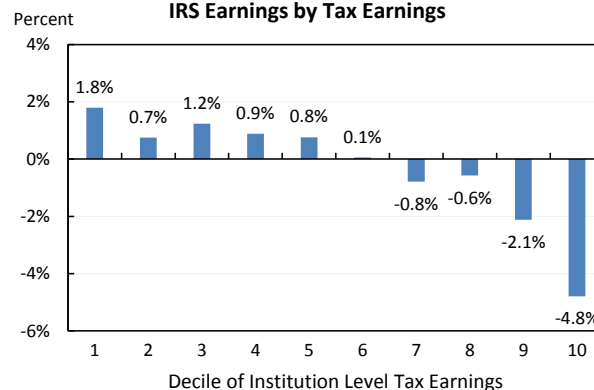

Source: Department of Treasury

Figure A-2b: Institutional Level Difference of UI Earnings Minus Tax Earnings by UI Coverage
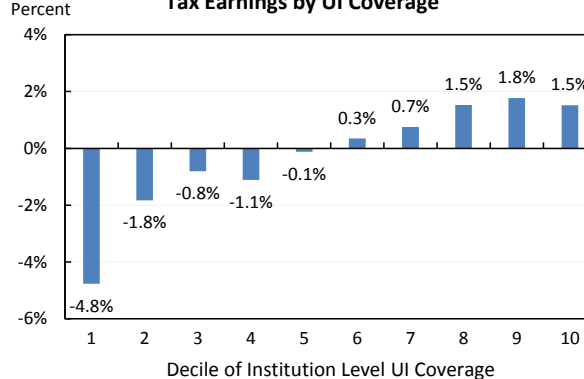
Source: Department of Treasury

Figure A-2c: Institutional Level Difference of UI Earnings Minus IRS Earnings by Tax Earnings

Source: Department of Treasury

On a whole, the IRS administrative tax earnings data used in the report provide some key improvements over the alternative UI measure. In particular, they offer a consistent measure of earnings-based outcomes for the universe of all Title IV students at all institutions of higher education. That said, the analyses above show that state efforts based on UI data are probably providing very worthwhile information, though in some cases there might be economically meaningful discrepancies. Even so, the comprehensiveness and consistency in these data allow for more transparency in outcomes than was previously available.

Completion

This section details the newly developed NSLDS completion data, that hold the promise of being able to measure outcomes for a broader set of students—including part-time and transfer students—and to provide more granular outcome estimates for various student subgroups. While NSLDS data is expected to improve over time, there are several limitations and concerns over the data quality behind the NSLDS rates that recommend caution in their use, especially for 'consumer information' applications. These data are being made available to the field to facilitate dialog about how these new data, in conjunction with other existing information, can best enhance estimates of institutional progression and completion related outcomes. Below, we discuss the current limitations of the NSLDS data through comparisons to three external data sources for completion: the National Student Clearinghouse (NSC), the Beginning Postsecondary Students Survey (BPS) and the State Council of Higher Education for Virginia (SCHEV). These representative data sources cannot provide easily accessible information at a national institutional level, but they are well-suited for comparing with the NSLDS data.

An overview of NSLDS completion data, including some features and limitations, is presented in Section III of the report. These NSLDS analyses show some promising trends in allowing future users to access reasonable data disaggregated by family income and other characteristics. For students with loans, NSLDS completion measures tend to look similar to other data sources, particularly in more recent years. However, reported data for students who only receive grants remains challenging to find, but progress is improving. As result, the data are not currently

recommended for consumer information applications, as they are not sufficiently accurate in many cases, but they are promising for the future.

Due to the issues described in Section III of the main report about weaknesses in NSLDS coverage of students who only receive Pell grants and do not take out federal loans, NSLDS completion rates were supplemented with NSC data for Pell-only recipients. In the 2007-2008 award year, roughly 2,600 (out of 5,941) institutions had a higher completion rate when including full supplemental NSC data. Of those institutions, the average difference between the two rates was roughly 2 percentage points with a standard deviation of around 2.7 percentage points, though at a handful of schools this difference was significantly larger. These statistics and Figure A-3 suggest that improved Pell-only reporting could more accurately capture institutional completion rates. The 2005-2006 and 2006-2007 cohorts had negligible differences, suggesting that NSC reporting may not have been as complete prior to the 2007-2008 award year.

Figure A-3: Completion Rates Pre- and Post-NSC Match at 4 Year Colleges
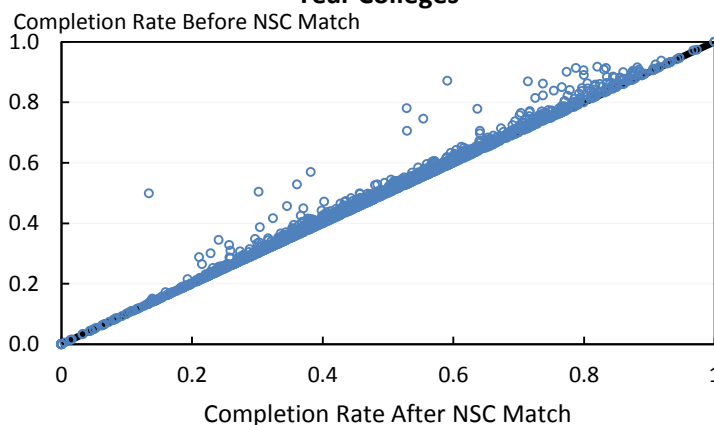
Note: Data are for the 2008 cohort. Cohorts less than 10 were dropped.
Source: Department of Education, College Scorecard Data

Comparisons of NSLDS data (supplemented by NSC data and including data for Pell-only students without an enrollment table record) with the closest cohort of BPS data provide national- and sector-level comparisons. The BPS surveys representative cohorts of first-time, beginning students at three points in time: the end of their first year as well as three and six years after first starting a postsecondary education. The BPS collects data on a variety of topics suitable for constructing a completion rate comparison. Table A-1 shows that the 2004 NSLDS national Title IV student six-year graduation rate was 39.5 percent, as compared to the six-year BPS Title IV graduation rate of 50.7 percent.[40] There was a slight upward trend in NSLDS completion rates from 2004 to 2008 based on limited NSC Pell-only completion rates. However, the 2007 and 2008 six-year completion rates with full Pell-only NSC supplemented data are still below the 2003-2004 BPS six-year completion rate. In addition, Table A-1 shows that public institutions have lower six-year completion rates in comparison to six-year BPS completion rates (though private for-profit

---

[40] The Title IV graduation rate is for students receiving Title IV aid in 2003-2004.

schools have higher six-year completion rates than six-year BPS completion rates Since Pell-only students are more prevalent in public institutions, Table A-1 suggests that even with full NSC supplemental data, NSLDS completion rates may underreport Pell-only completions.

Table A-1: Six-Year NSLDS and Five-Year BPS Title IV Completion Rates by Sector							
NSLDS sector	2004 cohort with no Pell-only NSC data	2007 cohort limited Pell-only NSC data	2008 cohort limited Pell-only NSC data	2007 cohort with Pell-only NSC data	2008 cohort with Pell-only NSC data	BPS 2004 cohort	BPS sector
All	39.5%	42.1%	43.5%	44.8%	47.4%	50.7%	All
Public 2-3 Year	17.7%	21.3%	23.5%	24.5%	29.3%	34.1%	Public 2 Year
Public 4+ Year	46.2%	49.8%	52.0%	52.7%	56.6%	61.7%	Public 4 Year
Private 2-3 Year	41.0%	46.1%	48.2%	49.1%	51.8%	45.8%	Private Not-for-profit 2 Year
Private 4+ Year	53.3%	55.3%	57.2%	58.0%	60.8%	67.7%	Private Not-for-profit 4 Year
Proprietary 2-3 Year	53.9%	56.1%	57.2%	57.4%	58.5%	39.1%	Private For-profit 2 Year
Proprietary 4+ Year	37.4%	36.7%	35.5%	38.9%	37.7%	32.9%	Private For-profit 4 Year
Source: NSLDS, BPS							

In a second exercise, NSLDS completion rates were supplemented with NSC data and compared to data provided by SCHEV, with a sample representing roughly 90 percent of the NSLDS observations. In order to examine the cohort construction in the NSLDS completion rate, a SCHEV variable describing the first time a student enrolled at an institution was compared with the first-aided methodology as described in the NSLDS completion rate methodology.

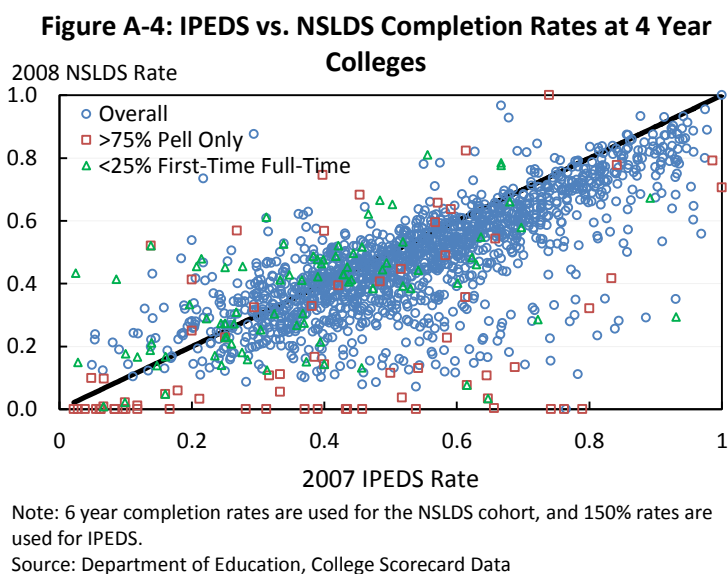| Table A-2: Year that Students Classified as a First Year in NSLDS Appear as a First Year in SCHEV				
	NSLDS 2008 cohort		NSLDS 2007 cohort	
Award year	Observations	Percent	Observations	Percent
Prior to 2001-02	2,632	3.96	2,831	5.21
2001-2002	741	1.12	840	1.55
2002-2003	888	1.34	1,096	2.02
2003-2004	1,264	1.9	1,547	2.85
2004-2005	1,861	2.8	2,483	4.57
2005-2006	2,670	4.02	4,004	7.37
2006-2007	4,808	7.24	**37,272**	**68.62**
2007-2008	**46,682**	**70.25**	2,348	4.32
2008-2009	2,567	3.86	1,334	2.46
2009-2010	1,784	2.68	207	0.38
2010-2011 and after	555	0.82	351	0.65
Source: NSLDS, SCHEV				

These data correspond reasonably well, but the analysis suggests that NSLDS completion reporting has room for improvement. For the sample in the 2007-2008 cohort, 70 percent of the observations identified in NSLDS were also identified as beginning students at an institution by SCHEV. In roughly 22 percent of the observations, SCHEV recorded a student starting at an institution prior to the 2007-2008 award year. Many of these observations may be due to students who first enrolled at an institution but did not receive Title IV aid at the time. Roughly 7 percent of observations in the NSLDS for this cohort were identified by SCHEV as students having started at the institution later than the cohort year. This is likely related to changes in the way SCHEV schools reported Social Security numbers and also suggests there may be some general inconsistencies between the two data systems. The 2006-2007 cohort has similar findings.

Observations with a completion event at a SCHEV school were also compared to NSLDS corresponding completion events in the same time period to assess the NSLDS completion measure. For each cohort, the time period spans between the start of the cohort year and the end of the 2013-14 award year (seven-year completion rate for 2007 cohort and eight-year completion rate for 2008 cohort). 23.1 percent of observations with a SCHEV completion event do not have a corresponding NSLDS completion event in the 2007 cohort. In the 2008 cohort, 16.8 percent of observations with a SCHEV completion event do not have a corresponding NSLDS completion event.

Table A-3: SCHEV Completion vs. NSLDS Completion by Sector					
Cohort	Pell Only	NSLDS Sector	Observations with SCHEV completion and without NSLDS completion	Observations with both SCHEV and NSLDS completion event	Percent with SCHEV completion without NSLDS completion
2007	No	Private 4+ Years	1,113	5,627	16.5%
		Public 2-3 Years	1,265	3,151	28.6%
		Public 4+ Years	1,984	14,614	12.0%
	Yes	Private 4+ Years	232	44	84.1%
		Public 2-3 Years	1,860	287	86.6%
		Public 4+ Years	731	139	84.0%
2008	No	Private 4+ Years	1,084	7,461	12.7%
		Public 2-3 Years	1,256	3,907	24.3%
		Public 4+ Years	2,043	17,163	10.6%
	Yes	Private 4+ Years	198	203	49.4%
		Public 2-3 Years	1,060	1,388	43.3%
		Public 4+ Years	539	574	48.4%
Source: NSLDS, SCHEV					

Table A-3 shows that SCHEV completion events occurred with greater frequency than NSLDS events for Pell-only observations. Around 85 percent of Pell-only students in the 2007 cohort who completed at a SCHEV school within six years did not have an NSLDS completion event over that timeframe. This improved measurably for the 2008 cohort, with only 45 percent of SCHEV completers not having an NSLDS completion event. These fractions are sizeable but decreasing. A final comparison between the NSC corrected 2008 NSLDS completion rate to the closest IPEDS rate shows that the two are reasonably well correlated at four-year schools (0.74). Schools with

a low fraction (less than a quarter) of first-time full-time students tend to be less correlated (0.50); this would be expected since the IPEDS measure is restricted to this small sample of students while the NSLDS measure is not. Additionally, schools with a high fraction (greater than three-quarters) of Pell-only students are less correlated due to errors in NSLDS data (0.50). Correlations at two-year schools are slightly lower (0.68).

Figure A-4: IPEDS vs. NSLDS Completion Rates at 4 Year Colleges

Note: 6 year completion rates are used for the NSLDS cohort, and 150% rates are used for IPEDS.
Source: Department of Education, College Scorecard Data

Comparisons with BPS and SCHEV data show that NSLDS data is not a perfect resource for understanding completion, though recent data show that there are some signs of improvement.

Cost

Understanding the costs of higher education is critical to both consumers, choosing whether and where to attend college, and to policymakers, examining whether scarce resources are being deployed in an efficient manner. Complications arise, however, because there are a variety of different ways to measure costs and the price students pay often differs from the `sticker price' they see. Various cost measures are outlined below, along with some less transparent costs that students and policymakers often consider, including opportunity costs. It is also worthwhile for interested parties to consider the ways in which institutional spending and cost relate.

- Tuition, sometimes called the `sticker price,' is what institutions generally advertise as the price charged to attend. The tuition measure reported in the College Scorecard is the sum of tuition and fees for full-time students, reported by institutions in IPEDS. Tuition and fees can vary by residency status, as many public academic-year institutions report costs separately for in-state and out-of-state students. Some colleges also have a separate tuition and fee schedule for 'in-district' students who reside in the same locality as the institution. Many other schools, especially private colleges, have the same tuition and fees for all students, regardless of residency status. For comparison purposes, Table A-4 below uses the in-state tuition measure when reported. For institutions on a non-

academic year that report tuition and fees by program year, it uses the values for tuition and fees that cover the full cost of the largest program.

- The broadest measure of costs to students is cost of attendance, also reported to IPEDS by institutions for students paying the in-state or in-district tuition rate.[41] In the cost of attendance measure, living expenses are calculated by weighting the expenses faced by students living on-campus, students living off-campus with their families, and students living off-campus without families by the number of students in each category. Though this measure provides students the most comprehensive view of the total cost of attending college (assuming they receive no additional aid), research by Kelchen, Hosch, and Goldrick-Rab (2014) find considerable variability in the accuracy of reported living expenses. They find that public institutions tend to underestimate actual living costs for students, while for-profits tend to provide the most inaccurate estimate, with both over and underestimation of expenses. Such inaccuracies can be costly for students deciding if and where to attend college.

- A final cost measure is the average net price to students at each institution, defined as the cost of attendance less all grant aid for federal aid recipients paying the in-state or in-district tuition rate. The net price measure captures the fact that many students who do not need to pay the full cost of attendance, as they receive grant aid from the state or federal government or the institution itself.

Table A-4 below describes the range of these various cost estimates across institutions in different sectors of higher education.

Table A-4: Overview of College Cost Measures										
		Public			Private Non-Profit			Private For-Profit		
School Type	Cost Measure	Median	10th Percentile	90th Percentile	Median	10th Percentile	90th Percentile	Median	10th Percentile	90th Percentile
< 2 Year	Tuition	3900	1900	12200	12900	5800	24900	15400	10300	20400
< 2 Year	Cost of Attendance	12900	8900	19800	19300	12300	32000	19600	12700	28400
< 2 Year	Net Price	8300	3600	16300	15000	6500	27900	15900	8600	24500
2 Year	Tuition	3400	1400	5300	16000	7900	26000	17100	11500	24000
2 Year	Cost of Attendance	12400	9300	16000	25300	15000	35500	26500	21500	30900
2 Year	Net Price	7500	4300	11500	19600	10500	27900	21900	15300	26100
4 Year	Tuition	7800	5800	13300	27000	10800	42800	16000	10600	21100
4 Year	Cost of Attendance	20100	15500	25500	38500	21100	57200	26300	23200	34800
4 Year	Net Price	14100	9100	18500	21500	12700	31600	20900	17900	30600
Note: Cost measures in this table are defined as follows. Tuition is the sticker price of the school. In-state prices are used for public schools when in-state and out-of-state prices differ. Cost of attendance is the noted tuition plus fees and supplies along with living expenses. Net price is the cost of attendance less grant aid.										
Source: Department of Education, College Scorecard Dara										

---

[41] There is some debate among education researchers as to whether living costs should be included in the cost of attendance, as estimates of living costs may not represent the additional cost relative to what an individual would need to spend if they were not in college (Johnson, 2009;Romano, Losinger, and Millard, 2011;College Board, 2014a).

While the table above shows that the net prices paid by students can differ substantially from the full cost of attendance on average, these data mask even greater differences for students from different family income backgrounds. On average, the poorest federal aid recipients pay a net price over $3,000 dollars less than those whose families earn over $30,000 a year, and $5,000 dollars less than those earning over $75,000 a year. The difference is driven in part by need-based federal and state financial aid, but also by institutions' financial aid policies. A recent study found, for example, that students paying the full tuition price are increasingly subsidizing poorer students through institutional grants (Delta Cost Project, 2010). Providing these net price data by family income categories is important in order to give students the most accurate data possible.

Figure A-5: Relationship Between Subsidization and Percent of Poor Students at 4 Year Schools

Difference Between Price Paid by Poorest Students vs. Other Students
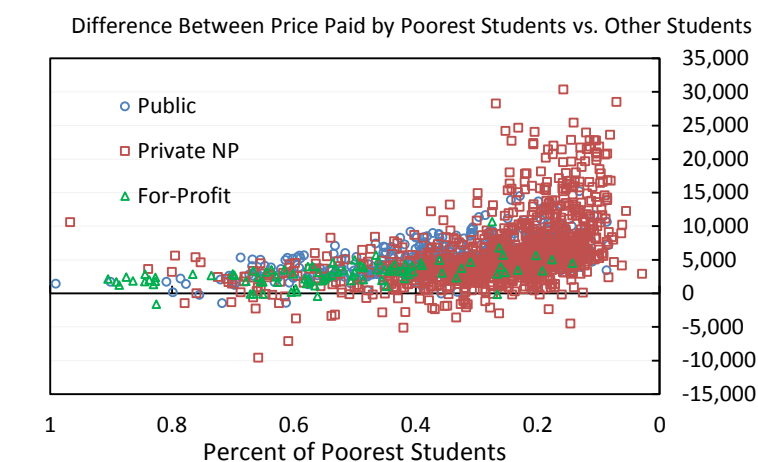
Source: Department of Education, College Scorecard Data

From an accountability perspective, it may be desirable to recognize institutions that target their own financial aid resources towards reducing net prices for the poorest students. A challenge in doing so, however, is that institutions may only be able to cross-subsidize the tuition of poorer students if there are relatively few poor students, since there must be a pool of students paying higher tuition to make such a strategy possible. Figure A-5 confirms that the gap between the net price charged to students with family incomes less than $30,000 and that charged to other students is greatest at institutions with the fewest poor students. As a result, a policy that incentivizes reducing the average net price charged to lower-income students may have the perverse effect of inducing institutions to enroll fewer low-income students.

Further, the cost schools incur in educating students can differ dramatically from the price students pay. Although many great schools can use their endowments to reduce costs to students, other schools may be spending in unsustainable ways, relying on external grant money from the government and other sources to fund their endeavors. About 500 private schools spend less than students' paid tuition on educational activities, instead saving some of the tuition as profit or spending a larger fraction of the tuition on buildings, athletics, or other expenses not

covered in the educational expense measure used. The most pervasive differences between net tuition and educational spending occur at private for-profit two-year schools.

Additionally, there is significant variation in the educational expenses at a school. Many schools rely on public grants, endowments, investments, and school operations to fund educational activities, with 42 schools spending more than 100,000 dollars per FTE on education. Curbing overall spending on all types of expenses is a priority for some institutions: as state and local appropriations are decreasing, schools will need to replace those funds through other means, including higher tuition (Kirshstein and Hurlburt, 2012).

Debt and Repayment

As students increasingly finance their education with loans, there is growing concern that a) financing costs unduly add to the already rising cost of college, b) some individuals may struggle to repay their loans after college and the resulting damage to credit records may hinder the ability to build wealth, and c) the burden associated with repaying loans may affect student choices after they exit college in undesirable ways.

Data from the National Center for Education Statistics' National Postsecondary Student Aid Survey (NPSAS), a sample of all undergraduates in 2011-2012, shows that the fraction of undergraduates with federal loans has increased from 43.1 percent in 2003-2004 to 52.3 percent in 2011-2012.[42] It is increasingly important, therefore, to provide students with information about how much they might need to borrow to attend various institutions and how likely they will be able to repay those loans once they leave college.

**Figure A-6: Change in Student Debt from 2000 to 2013**
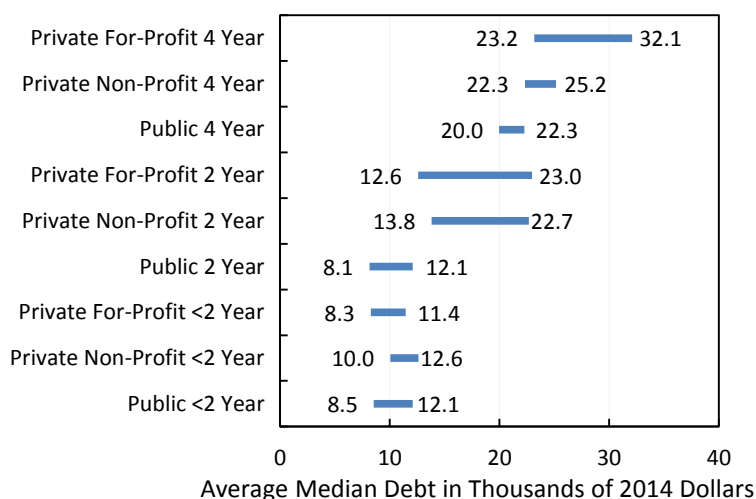
Source: Department of Education, College Scorecard Data

---

[42] A New America policy brief (Miller, 2014) calculates a similar percentage but also includes private loans, finding an increase from 53.5 to 61.6 percent from 2003-2004 to 2011-2012 for completers.

It is important to note that this measure does not reflect the amount borrowed by the typical (i.e., median) student at an institution, since the fraction of students who borrow varies greatly across institutions, as shown in Figure A-7 below. For example, a majority of students at nearly all two-year for-profits borrow, while a minority of students do so for nearly all two-year public schools. At four-year schools, borrowing is common across public, private non-profit, and for-profit institutions, but, again, is most prevalent at for-profits and least prevalent at public institutions.

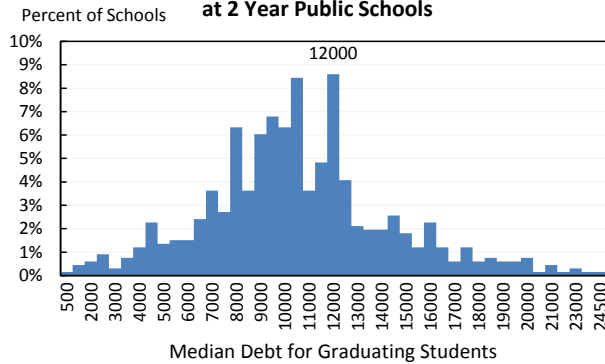**Figure A-7: Percent of Students with Federal Loans by Sector**
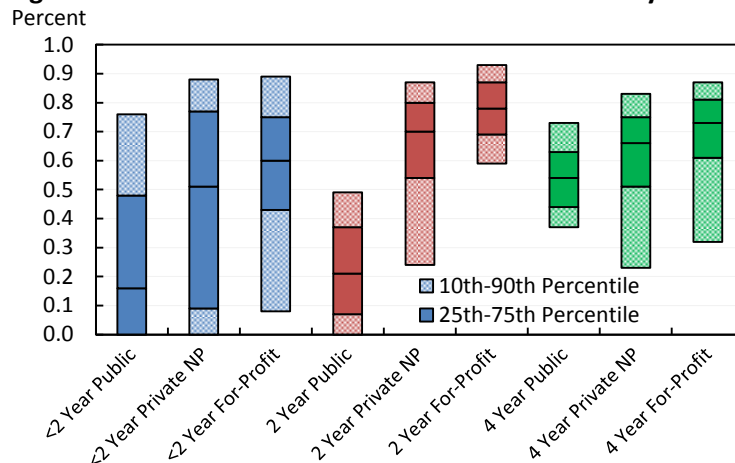
Note: Data are for the fraction of all undergraduates with a federal loan in 2013.
Source: Department of Education, College Scorecard Data

Additionally, in the past, the initial College Scorecard reported the median debt for all students who borrow at an institution. This reflects the amount that the typical student who is expecting to borrow will borrow. However, at institutions where a large number of students withdraw before completion, the total debt level is likely to be influenced by the fact that the typical student spends only a fraction of the time it takes to complete a degree at the institution. Using the median debt level for students who complete a degree avoids this distortion; this measure is used throughout the report since it is more likely to reflect affordability differences across institutions. It is important to note, however, that this measure (appropriately) is still influenced by time-to-degree differences across institutions: institutions where students take longer to complete are likely to have higher debt levels.
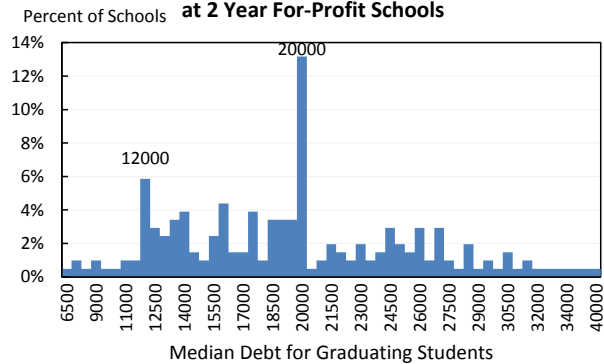
A closer look at the median cumulative debt measure shows that typical students at many institutions—especially four-year private non-profit institutions—have total borrowing amounts equal to federal loan limits. As shown in Figures A-8a-d, there are large spikes in the distribution of total borrowing that correspond with the Direct Loan limit for dependent undergraduates at each type of institution: students can borrow up to $12,000 for two years at two-year schools and up to $27,000 for four years of college at four-year schools. Independent undergraduates are capped at $20,000 for two years of college.[43]

---

[43] Loan limits are defined for annual borrowing and differ by grade level, but they sum to the amounts noted.
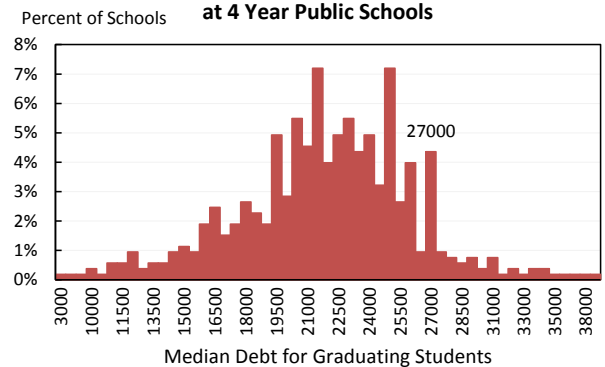
Figure A-8a: Distribution of Debt for Graduating Students at 2 Year Public Schools
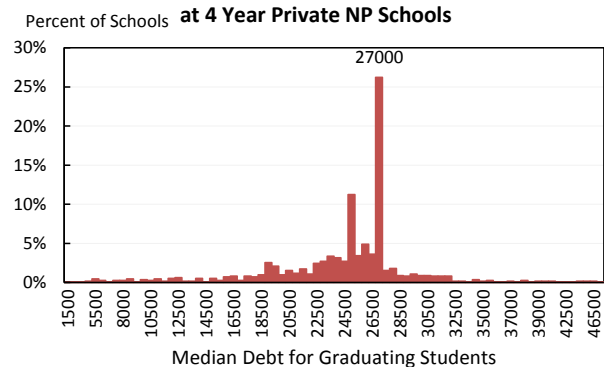
Percent of Schools

Median Debt for Graduating Students

Note: Data are for the 2013 and 2014 exit cohorts
Source: Department of Education, College Scorecard Data

Figure A-8b: Distribution of Debt for Graduating Students at 2 Year For-Profit Schools

Percent of Schools

Median Debt for Graduating Students

Note: Data are for the 2013 and 2014 exit cohorts
Source: Department of Education, College Scorecard Data

Figure A-8c: Distribution of Debt for Graduating Students at 4 Year Public Schools

Percent of Schools

Median Debt for Graduating Students

Note: Data are for the 2013 and 2014 exit cohorts
Source: Department of Education, College Scorecard Data

Figure A-8d: Distribution of Debt for Graduating Students at 4 Year Private NP Schools

Percent of Schools

Median Debt for Graduating Students

Note: Data are for the 2013 and 2014 exit cohorts
Source: Department of Education, College Scorecard Data

Existence of these caps leads to a break in the relationship between debt and net price, which tends to be most binding in the private non-profit sector. The fact that borrowing is at or near the federal limit for the typical student at so many colleges highlights another limitation with the NSLDS data: they exclude several types of debt. Parent PLUS loans are not included. NPSAS data show that, although a small minority of students' parents borrow with these loans (at about 4.5 percent overall), they tend to be sizeable, with a median of about 10,000 dollars among PLUS borrowers (NPSAS, 2012). The same dataset shows that only 6 percent of borrowers hold any private loans, and for those borrowers, median debt amounts tend to be small, at 4,000 dollars (NPSAS, 2012). The shares of parent and private borrowers tend to be concentrated at four-year and less-than-two-year schools. These data may understate the extent of private borrowing for education, however, since families might also make other financial adjustments to pay for college that will not be captured under the student debt measure, such as taking a home equity loan. While the cumulative debt measure used cannot capture all of the financial burdens of enrolling in a particular school – especially some private loans that could be more costly – it likely captures most debt taken on by students and their families to attend college (College Board, 2014b; TICAS, 2014).

Figure A-9: Relationship Between Debt and Net Price at 4 Year Schools

Median Graduate Debt
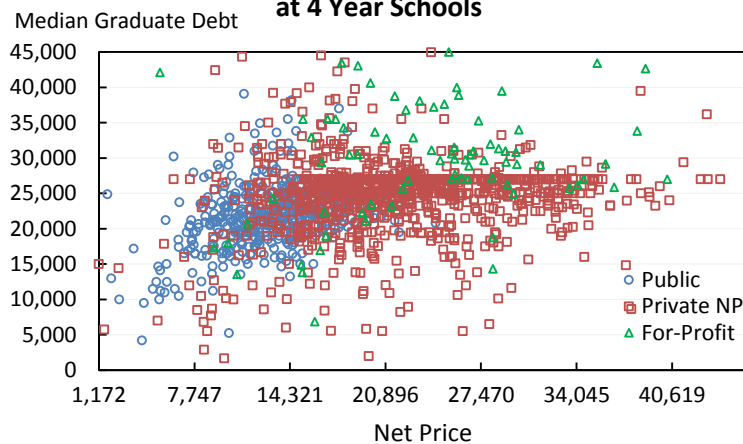
Note: Data are for the 2013-2014 pooled debt cohort and 2013 net price.
Source: Department of Education, College Scorecard Data

Another critical consideration for consumers and policymakers is the repayment rate, which shows the fraction of students at an institution who are able to pay down their loans. This measure is similar to the well-known cohort default rate metric (CDR) but is meant to be less susceptible to gaming behavior by institutions.[44] In addition to default, individuals count as a 'failure' if their loan balance fails to decline between repayment entry and the measurement date. As such, forbearance or deferments alone do not improve the institutions' rate since loan interest will still generally accrue even if this prevents a loan from entering default. One flaw with treating negative amortization of the loan as a failure is that students with unsubsidized loans who might defer for 'good reasons'—e.g., to attend graduate school—may see their loan balances increase between measurement periods as interest accumulates even if their loan is always in good standing. This problem is mitigated somewhat by excluding students in in-school loan deferment at the point of measurement from the calculation. But the measure would still record as a 'failure,' for example, a student who entered repayment, saw their balance increase while in a master's program, and then entered repayment again just prior to the three-year measurement date.

The overall three-year repayment rate for all undergraduate institutions, weighted by the number of students borrowing at each school, was 63 percent in the combined 2010 and 2011 repayment cohorts. 37 percent of students were thus not meeting the repayment metric—either they were in default or were making monthly payments that were not reducing their loan balance. As a point of comparison, the three-year CDR was 13 percent for all students in the 2011 repayment cohort.

It is expected that the default rate is lower than the repayment rate, since the repayment rate captures not only those loans in default, but also those that may potentially be in deferment or

---

[44] From the late 1990s through the 2000s, an increased length of the delinquency period of a loan (180 to 270 days), coupled with a measurement period of 2 years, better allowed for deferment and forbearance periods to mask student default, leading to significantly lower default rates (Lederman, 2007).

forbearance, or for other reasons carry a balance exceeding the balance of the loan at repayment. An important case where this might happen is if a borrower enrolls in an income-driven repayment plan and, despite making on-time payments, may not actually be making progress towards lowering the balance of the loan. This could be because payments only cover the interest of the loan or because payments due are zero for those with low enough income.

Compared to the cohort default rate, the repayment rate measure is designed to be a more robust measure of the fraction of borrowers struggling to repay their loans. In particular, it cannot be improved by nudging students into forbearance or deferment, as that alone will not lead a student's loan balance to decline. Interest will also continue to accumulate during forbearance on all loans and during deferment for unsubsidized loans. As expected, the 'non-repayment rate' (i.e., 1 minus the repayment rate) is consistently larger than the cohort default rate. The difference between the cohort default rate and the 'non-repayment rate' is largest at two-year and less-than-two-year colleges, where the average difference is roughly 25 percentage points, as compared to 12.5 percentage points at four-year colleges. Differences are driven by for-profit institutions, where the average difference is about 30 percentage points, followed by public schools and private non-profits. These differences are important to note, as we believe that repayment rates more accurately reflect the borrowing behavior of students than default rates. Temporal features of the repayment rate metric are important as well. Rates are stable across various time horizons, especially between measures at least three years after repayment. For example, the repayment rates for the 2007 cohort, the most recent for which all longer-term measures are available, show an overall student-level repayment rate of 81 percent at year one and 75 percent for all of years three, five, and seven. However, differences across cohorts tend to be less stable. Three-year repayment rates have steadily dropped from 75 percent in 2007 to 61 percent in 2011, reflecting a variety of factors, including a shift of borrowers to the for-profit sector where repayment outcomes tend to be worse and a decline in labor market opportunities for all students (Looney and Yannellis, 2015).

Access

Improving outcomes and reducing costs is especially important for disadvantaged students. College Scorecard data show that advantaged students often have better completion rates, labor market outcomes, and loan repayment rates than disadvantaged students who attend the same colleges, indicating that colleges face a particular challenge in serving low-income students who enroll.[45]

In order to understand and improve higher education opportunities for disadvantaged students, it is first necessary to measure enrollment of these students at institutions. In this section, we outline four ways that enrollment of disadvantaged students in higher education could be measured. First, IPEDS data can be used to determine the fraction of Pell students at an

---

[45] As mentioned above, there are exceptions to this general rule. For instance, student characteristics exhibit a negligible relationship with the gainful employment metric (debt to earnings ratio) among applicable programs; this is much less the case when comparing a broader set of outcomes across the full spectrum of higher education institutions.

institution. This measure is simple to calculate but obscures differences in financial circumstances among students who receive Pell grants and those who do not and miscounts the fraction of low income students in places where low tuition and alternative grants reduce the need for Pell.

IPEDS data can also be used for measuring enrollment of disadvantaged students by examining the fraction of students in different family income bands, where low-income status may be reasonably defined as students with family income less than $48,000. Using the IPEDS data to construct a family income measure requires restricting analysis to first-time full-time students who receive Title IV aid, while the percentage Pell measure relates to all students at an institution. These measures are correlated (0.74 at four-year schools), as shown below in Table A-5, but differences may arise because low income enrollment rates will exclude part-time, transfer and non-Title IV students. Similar NSLDS data can also be used to calculate the fraction of low-income enrollment.

Alternatively, parental financial and parental education data from the NSLDS can be used to construct enrollment rates of students with different demographic backgrounds. Though this data is only available for Title IV students, the demographic information in the NSLDS can be used to calculate the fraction of first-generation students whose parents did not attend college.
A final enrollment measure could be defined by the poverty rate in the students' zip code using U.S. Census data merged to individual student records in the NSLDS. As with other NSLDS calculations, this enrollment rate is restricted to Title IV students which may decrease the comprehensiveness of the measure. Correlations at two- and four-year schools for these measures are noted below.

Table A-5: Correlation Across Access Measures									
2 Year					4 Year				
	Fraction Pell	Fraction Low Income	First Generation Student	Zip Poverty Rate		Fraction Pell	Fraction Low Income	First Generation Student	Zip Poverty Rate
Fraction Pell	1.00				Fraction Pell	1.00			
Fraction Low Income	0.20	1.00			Fraction Low Income	0.74	1.00		
First Generation Student	0.32	0.30	1.00		First Generation Student	0.58	0.51	1.00	
Zip Poverty Rate	0.30	0.41	0.19	1.00	Zip Poverty Rate	0.61	0.65	0.43	1.00
Note: Measures are identical to those used in Figures 1, 2, and 3.									
Source: Department of Education, College Scorecard Data									

Colleges face a unique challenge in serving disadvantaged students. The measures described above can be a useful starting point to examine how well colleges serve these students.

References

Altonji, Joseph and Richard Mansfield. 2014. "Group-Average Observables as Controls for Sorting on Unobservables When Estimating Group Treatment Effects: the Case of School and Neighborhood Effects." NBER Working Paper No. 20781.

Avery, Christopher and Thomas Kane. 2004. "Student Perceptions of College Opportunities. The Boston COACH Program." In *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It*: 355-394. Edited by Caroline M. Hoxby. University of Chicago Press.

Baum, Sandy, Jennifer Ma, and Kathleen Payea. 2013 "Education Pays: The Benefits of Higher Education for Individuals and Society." CollegeBoard: Trends in Higher Education Series.

Baumol, William J. 1990. "Entrepreneurship: Productive, Unproductive, and Destructive." *Journal of Political Economy* 98(5): 893-921.

Bettinger, Eric P. and Bridget Terry Long. 2010. "Does Cheaper Mean Better? The Impact of Using Adjunct Instructors on Student Outcomes." *The Review of Economics and Statistics* 92(3): 598-613.

Bettinger, Eric P. et al. 2009. "The Role of Simplification and Information in College Decisions: Results from the H&R Block FAFSA Experiment." NBER Working Paper No. 15361.

Betts, John. 1996. "What Do Students Know About Wages? Evidence from a Survey of Undergraduates." *The Journal of Human Resources* 31(1): 27-56.

Black, Dan and Jeffrey Smith. 2006. "Evaluating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics* 24(3): 701-728.

Carnevale, Anthony P., Stephen J. Rose, and Ban Chea. 2014. "The College Payoff: Education, Occupations, Lifetime Earnings." Georgetown University Center on Education and the Workforce.

Chetty, Raj, et al. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review* 104(5): 141-47.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014 "Measuring the Impact of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104(9): 2633-2679.

Cohodes, Sarah and Joshua Goodman. 2012. "First Degree Earns: The Impact of College Quality on College Completion Rates." Harvard Kennedy School Faculty Research Working Paper Series RWP12-033.

College Board. 2010. "Student Poll: Students and Parents Making Judgments about College Costs without Complete Information." *Student Poll 8(1).*

College Board. 2014a. "Trends in College Pricing: 2014." *Trends in Higher Education.*

College Board. 2014b. "Trends in Student Aid: 2014." *Trends in Higher Education.*

College Measures. 2015. "Economic Success Measures—Texas." Retrieved from http://esm.collegemeasures.org/esm/texas/default.aspx/.

College Results Online. 2015 "How College Results Online Identifies "Similar" Institutions." The Education Trust. Retrieved from http://collegeresults.org/aboutthedata.aspx#section-5.

Cunha, Jesse M. and Trey Miller. 2014. "Measuring value-added in higher education: Possibilities and limitations in the use of administrative data." *Economics of Education Review* 42: 64-77.

Dale, Stacy Berg and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics* 107(4): 1491-1527.

Dale, Stacy Berg and Alan B. Krueger. 2011. "Estimating the Return to College Selectivity over the Career Using Administrative Earnings Data." NBER Working Paper No. 17159.

Delta Cost Project. 2010. "Who Pays for Higher Education? Changing Patterns in Cost, Price, and Subsidies." American Institutes for Research.

Department of Education. 2015. "Fact Sheet: Obama Administration Increases Accountability for Low-Performing For-Profit Institutions." Retrieved from http://www.ed.gov/news/press-releases/fact-sheet-obama-administration-increases-accountability-low-performing-profit-institutions.

Desrochers, Donna M. and Jane V. Wellman. 2011. "Trends in College Spending: 1999-2009." Delta Cost Project: American Institutes for Research.

Dominitz, Jeff and Charles F. Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources* 31: 1-25.

Eagan, Kevin et al. 2014. "The American Freshman: National Norms Fall 2014." Cooperative Institutional Research Program at UCLA.

Fishman, Rachel. 2015. "2015 College Decisions Survey Part 1: Deciding to Go to College." New America Education Policy Program.

Goodman, Joshua, Michael Hurwitz, and Jonathan Smith. 2015. "College Access, Initial College Choice and Degree Completion." NBER Working Paper No. 20996.

Grodsky, Eric and Melanie Jones. 2007. "Real and Imagined Barriers to College Entry: Perceptions of Cost." *Social Science Research* 36: 745-766.

Hanushek, Eric and Margaret Raymond. 2005. "Does School Accountability Lead to Improved School Performance?" *Journal of Policy Analysis and Management* 24(2): 297-329.

Haskins, Ron. 2008. "Education and Economic Mobility." The Brookings Institution.

Hastings, Justine S. et al. 2015. "(Un)informed College and Major Choice: Evidence from Linked Survey and Administrative Data." NBER Working Paper No. 21330.

Heckman, James J., Lance J. Lochner, and Petra E. Todd. 2006. "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." *Handbook of the Economics of Education* 1: 307-458.

Hill, Kent, Dennis Hoffman, and Tom R. Rex. 2005. "The Value of Higher Education: Individual and Societal Benefits (With Special Consideration for the State of Arizona)." Arizona State University: Productivity and Prosperity Project.

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *The Review of Economics and Statistics* 91(4): 717-724.

Horn, Laura, Xianlei Chen, and Chris Chapman. 2003. "Getting Ready to Pay for College: What Students and Their Parents Know About the Cost of College Tuition and What They Are Doing to Find Out." U.S. Department of Education National Center for Education Statistics.

Hoxby, Caroline M. 2015. "The Value-Added of American Postsecondary Institutions."

Hoxby, Caroline and Sarah Turner. 2013. "Expanding College Opportunities for High-Achieving, Low Income Students." SIEPR Discussion Paper No. 12-014.

Institute of Medicine. 1990. *Medicare: A Strategy for Quality Assurance, Volume 1.* The National Academy Press.

Internal Revenue Service (IRS). 2014. "Wages, Salaries, and Other Earnings: Special Rules for Certain Employees." Chapter in Publication 17: Your Federal Income Tax. Retrieved from http://www.irs.gov/publications/p17/index.html.

Jacobson, Louis and Robert LaLonde. 2013. "Using Data to Improve the Performance of Workforce Training." The Hamilton Project and Results for America.

Jacobson, Louis, Robert LaLonde, and Daniel Sullivan. 2005. "The Impact of Community College Retraining on Older Displaced Workers: Should We Teach Old Dogs New Tricks?" *Industrial and Labor Relations Review* 58(3): 398-415.

Jaeger, David A. and Marianne E. Page. 1996. "Degrees Matter: New Evidence on Sheepskin Effects in the Return to Education." *The Review of Economics and Statistics* 78(4): 733-740.

Johnson, Nate. 2009. "What Does a College Degree Cost? Comparing approaches to measuring 'cost per degree.'" Delta Cost Project: White Paper Series.

Jones, Dennis P. 2013. "Outcomes-Based Funding: The Wave of Implementation." Complete College America and the National Center for Higher Education Management Systems.

Kane, Thomas J. and Cecilia Elena Rouse. 1995. "Labor-Market Returns to Two- and Four-Year College." *The American Economic Review* 85(3) 600-614.

Kane, Thomas J. and Douglas O. Staiger. 2002. "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives* 16(4): 91-114.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper No. 14607.

Kane, Thomas J. et al. 2013. "Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment." Bill and Melinda Gates Foundation: Measures of Effective Teaching Project.

Kelchen, Robert, Braden J. Hosch, and Sara Goldrick-Rab. 2014. "The Cost of College Attendance: Trends, Variation, and Accuracy in Institutional Living Cost Allowances." Wisconsin Hope Lab.

Kelchen, Robert and Douglas N. Harris. 2012. "Can 'Value Added' Methods Improve the Measurement of College Performance? Empirical Analyses and Policy Implications." HCM Strategists.

Kirshstein, Rita J. and Steven Hurlburt. 2012. "Revenues: Where Does the Money Come From?" Delta Cost Project: American Institutes for Research.

Lederman, Doug. 2007. "A More Meaningful Default Rate." *Inside Higher Ed.*

Lockwood, Benjamin B., Charles G. Nathanson, and E. Glen Weyl. 2012. "Taxation and the Allocation of Talent."

Long, Mark C. 2008. "College Quality and Early Adult Outcomes." *Economics of Education Review* 27(5): 588-602.

Looney, Adam and Constantine Yannellis. 2015. "A Crisis in Student Loans? How Changes in the Characteristics of Borrowers and in the Institutions They Attend Contributed to Rising Loan Defaults." *Brookings Papers on Economic Activity*.

McClellan, Mark and Douglas Staiger. 1999. "The Quality of Health Care Providers." NBER Working Paper No. 7327.

McClellan, Mark and Douglas Staiger. 2000. "Comparing the Quality of Health Care Providers." In *Frontiers in Health Policy Research*, 3: 113-136. Edited by Alan M. Garber. MIT Press.

Miller, Ben. 2014. "The Student Debt Review: Analyzing the State of Undergraduate Student Borrowing." New America: Education Policy Program.

National Conference of State Legislatures. 2015. "Performance-Based Funding for Higher Education." Retrieved from http://www.ncsl.org/research/education/performance-funding.aspx.

National Postsecondary Aid Survey (NPSAS). 2012. National Center for Education Statistics. Retrieved from https://nces.ed.gov/surveys/npsas/.

National Postsecondary Education Cooperative. 2007. "Deciding on Postsecondary Institution: Final Report." U.S. Department of Education National Center for Education Statistics.

OECD. 2013. "Education Indicators in Focus."

PayScale. 2015. "PayScale College ROI Report – 2015 Methodology and Notes." Retrieved from http://www.payscale.com/college-roi/methodology.

Radford, Alexandria Walton, Shayna Tasoff, and Thomas Weko. 2009. "Choosing a Postsecondary Institution: Considerations Reported by Students." U.S. Department of Education National Center for Education Statistics.

Romano, Richard M., Regina Losinger, and Tim Millard. 2011. "Measuring the Cost of a College Degree: A Case Study of a SUNY Community College." *Community College Review* 39(3): 211-234.

Rothstein, Jesse. 2010. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics* 125(1): 175-214.

Rothwell, Jonathan and Siddharth Kulkarni. 2015. "Beyond College Rankings: A Value-Added Approach to Assessing Two- and Four-Year Schools." Brookings: Metropolitan Policy Program.

Ruder, Alex and Michelle Van Noy. 2014. "The Influence of Labor Market Outcomes Data on Major Choice: Evidence from a Survey Experiment." Heldrich Center for Workforce Development.

Snyder, Martha. 2015. "Driving Better Outcomes: Typology and Principles to Inform Outcomes-Based Funding Models." HCM Strategists.

Tennessee Higher Education Commission. 2015. "Outcomes Based Funding Formula Resources." Retrieved from http://www.tn.gov/thec/topic/funding-formula-resources.

The Institute for College Access and Success (TICAS). 2014. "Quick Facts About Student Debt."

U.S. News and World Report. 2014. "How U.S. News Calculated the 2015 Best Colleges Rankings." Retrieved from http://www.usnews.com/education/best-colleges/articles/2014/09/08/how-us-news-calculated-the-2015-best-colleges-rankings.

Washington Monthly. 2014. "2014 National Universities Rankings." Retrieved from http://www.washingtonmonthly.com/college_guide/rankings-2014/national-universities-rank.php.

Wiswall, Matthew and Basit Zafar. 2013. "How Do College Students Respond to Public Information about Earnings?" Federal Reserve Bank of New York Staff Reports No. 516.

Zimmerman, Seth. 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics* 32(4): 711-754.