

NBER WORKING PAPER SERIES

LABOR MARKET OUTCOMES AND POSTSECONDARY ACCOUNTABILITY:
ARE IMPERFECT METRICS BETTER THAN NONE?

Veronica Minaya
Judith Scott-Clayton

Working Paper 22880
<http://www.nber.org/papers/w22880>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2016

Authors are listed in alphabetical order; both contributed equally to this project. The authors gratefully acknowledge Josh Hawley and Lisa Neilson at the Ohio Education Research Center, who helped facilitate our application for the restricted data utilized herein, and Caroline Hoxby, Kevin Stange, Jeff Smith, Robert Kelchen, and NBER conference participants for valuable suggestions. Funding to obtain and clean the data used herein was provided by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C110011 to Teachers College, Columbia University. The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, the Ohio Education Research Center, or the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Veronica Minaya and Judith Scott-Clayton. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Labor Market Outcomes and Postsecondary Accountability: Are Imperfect Metrics Better than None?

Veronica Minaya and Judith Scott-Clayton

NBER Working Paper No. 22880

December 2016

JEL No. H52,I23,I26

ABSTRACT

Policymakers at the state and federal level are increasingly pushing to hold institutions accountable for the labor market outcomes of their students. There is no consensus, however, on how such measures should be constructed, or how the choice of measure may affect the resulting institutional ratings. Using state administrative data that links postsecondary transcripts and in-state quarterly earnings and unemployment records over more than a decade, we construct a variety of possible institution-level labor market outcome metrics. We then explore how sensitive institutional ratings are to the choice of labor market metric, length of follow-up, and inclusion of adjustments for student characteristics. We also examine how labor market metrics compare to the academic-outcome-based metrics that are more commonly incorporated into state accountability systems. We conclude that labor market data provides information that is quite distinct from students' academic outcomes. Institutional ratings based on labor market outcomes, however, are quite sensitive to the specific metric. The choice of metric and length of follow up appear to matter much more than compositional adjustments. Our findings suggest a cautious approach: while a mix of feasible labor market metrics may be better than none, reliance on a single metric, especially if measured very early, may undermine policymakers' ongoing efforts to accurately quantify institutional performance.

Veronica Minaya
Teachers College
Columbia University
525 W. 120th St, Box 174
New York, NY 10027
vmm2122@tc.columbia.edu

Judith Scott-Clayton
Teachers College
Columbia University
525 W.120th Street, Box 174
New York, NY 10027
and NBER
scott-clayton@tc.columbia.edu

I. Introduction

The postsecondary accountability movement is motivated by the idea that reporting and rewarding measures of institutional performance can generate both better information and stronger financial incentives to improve the decision-making processes of prospective consumers, policymakers, and institutions (Dougherty & Reddy, 2013). In his 2013 State of the Union address, President Obama gave voice to this movement by calling for institutions to be “[held] accountable for cost, value, and quality,” eventually by linking measures of institutional performance to federal aid (U.S. Department of Education, 2013).

This accountability agenda is even more advanced at the state level. As of 2015, 32 states were already utilizing some form of performance or “outcomes-based” funding, with another 5 in the process of implementing it (National Council of State Legislatures [NCSL], 2015). While in most states the portion of state funding that is performance-based remains small—typically less than 10 percent—two states (Tennessee and Ohio) now base the majority of institutional funding on performance metrics (Snyder, 2015).

These accountability efforts increasingly look beyond just credit and credential completion to what some view as the most important dimension of student outcomes: post-college labor market success. In September 2015, the Obama administration took a major step towards this goal by releasing an updated version of its College Scorecard, which for the first time provided information not just on college costs and graduation rates, but also on median post-college earnings for over 4,000 institutions nationwide. Several states now incorporate job placement, employment and earnings data into their performance funding formulae, as least for portions of their postsecondary sectors. And the Texas State Technical College System uses information on students’ post-college earnings as the *sole* criteria for making funding

recommendations to the Texas legislature (Texas Higher Education Coordinating Board [THECB], 2013; Selingo & Van Der Werf, 2016).

There is no consensus, however, on how such labor market measures should be constructed, nor is there much evidence regarding how the choice of measure may affect the resulting institutional ratings. While the College Scorecard provides earnings for all entrants 10 years after entry, states using labor market data in performance funding formulae sometimes examine outcomes for graduates less than a year after graduation. Does it matter whether employment/earnings are measured one, two, or ten years post-graduation? Moreover, should schools be held accountable for all students, or just those who graduate? What difference does it make whether metrics are adjusted to account for the incoming characteristics of the student population? And can labor market data be used to examine more than just earnings?

In this paper, using administrative data from one state that links postsecondary transcripts to in-state quarterly earnings and unemployment records over more than a decade, we construct a variety of possible institution-level labor market outcome metrics. Our goal is not to identify the “best” metric, but to explore how sensitive institutional ratings may be to the choice of metric, length of follow-up, and inclusion of adjustments for student characteristics, particularly in the context of real-world data limitations. We believe we are the first to use a state-level database to assess labor market outcome metrics beyond earnings, including full-time, full-year employment rates, social service sector employment, and unemployment claims. We also examine how these metrics compare to the academic-outcome-based metrics that are more commonly incorporated into state accountability systems. This work builds upon similar efforts to analyze labor outcome metrics in the postsecondary sector using Internal Revenue Service data (Hoxby, 2015, 2016), the College Scorecard data (Executive Office of the President [EOP], 2015), using data on

four-year colleges in Texas and Canada (Cunha & Miller, 2014; Betts, Ferrall, & Finnie, 2013), . It also builds upon research on institutional performance measurement in sectors with similar features, including job training (Heckman, Heinrich, & Smith, 2002) and health care (Staiger, Chen, Birkmeyer, Ryan, Zhang, & Dimick; 2013).

We conclude that labor market data, even when imperfect, *can* provide valuable information distinct from students' academic outcomes. As has been found in other sectors, however, ratings are highly sensitive to the choice of outcome and length of follow up (and, to a lesser extent, to the inclusion of compositional adjustments). The most obvious labor market outcomes – graduates' employment and earnings in the year after graduation – are unreliable predictors of institutional performance on the same metrics measured several years later. Earnings and employment alone also fail to capture other aspects of economic wellbeing that may be valued by both policymakers and students themselves. Consistent with Cunha & Miller (2014), our findings suggest a cautious approach: while a mix of feasible labor market metrics may be better than none, reliance on a single unadjusted earnings metric, especially if measured too early, may undermine policymakers' ongoing efforts to accurately quantify institutional performance.

The remainder of the paper proceeds as follows. Section II provides policy context around performance accountability efforts in higher education. Section III provides a conceptual and practical overview of the challenges to using state-level labor market data for this purpose. Section IV describes our data and methodology. Section V presents results, and Section VI provides a concluding discussion.

II. Policy background

Policy goals in higher education traditionally have been measured and financed primarily using input metrics—such as student enrollment and credit hours—for many decades (SRI International, 2012).¹ This stands in contrast to the job training sector, which has a more established tradition of evaluating programs based on participants' labor market outcomes, going back at least to the Job Training Partnership Act of 1982 (Barnow & Smith, 2015). Over the past three decades, however, there has been a push to align higher education funding with academic outputs, such as credits completed or degrees conferred, rather than inputs. Output-based accountability efforts range from purely informational reporting to higher-stakes performance-based funding (Burke, 2001; Umbricht et al., 2015; Dougherty & Reddy, 2013). The idea behind outcomes-based accountability policies is that they generate both better information and stronger financial incentives to improve the decision-making processes of prospective consumers, policymakers, and institutions (Heckman, Heinrich, & Smith, 2002; Muriel & Smith, 2011; Dougherty et al., 2014).

The first wave of performance funding (PF) policies, or PF 1.0, used metrics to award bonuses over and above base state funding for higher education (Dougherty et al., 2012; Dougherty & Reddy, 2013; Snyder, 2011). These early programs eventually lost support, however, due to dissatisfaction with the reliability and validity of the chosen performance metrics, the top-down process by which they were determined, and the small amount of funding at stake (Snyder, 2015; Burke & Serban, 1997). More than half of PBF 1.0 programs were abandoned in the early 2000s (Dougherty et al., 2012).

¹ A cost-plus approach is a traditional budgeting strategy, in which public colleges and universities primarily based their projected budgetary needs on current costs, student enrollments and inflationary increases.

A new wave of performance funding, which “no longer takes the form of a bonus but rather is part and parcel of regular state base funding for higher education” (Dougherty et al., 2014) began spreading in the late 2000s. Ohio and Indiana both established such PF 2.0 programs in 2009, followed by Tennessee in 2010 (Dougherty et al., 2014; Dougherty & Reddy, 2013). By 2015, 32 states had a policy in place to allocate a portion of funding based on performance indicators, with 5 others in the process of transitioning (NCSL, 2015). Although many states continue to use performance funding to allocate relatively small percentages of higher education funding, some states now allocate much larger percentages of funding using performance metrics (Dougherty & Reddy, 2013). For example, outcomes-based funding represents about 2/3 of total state support to all higher education institutions (Snyder, 2015). This high proportion of funding is one reason why Snyder (2015) classifies Ohio and Tennessee as the two most advanced/high-stakes funding systems, that some are calling PF 3.0 (Kelchen & Stedrak, 2016).

With respect to the range of outcomes considered, 28 states currently 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 (NCSL, 2015). Many states give extra weight to outcomes for certain subgroups, such as Pell-eligible students (Burke, 2002; Dougherty, et al, 2009). Recently and particularly after the recession, accountability conversations have increasingly focused on the financial costs and benefits of college. Ten states now put weight on post-graduation outcomes such as job placement rates or earnings (EOP, 2015; NCSL, 2015; see Appendix Table A.1 for additional details). The Texas State Technical College System now uses information on students’ post-college earnings as the *sole* criteria for making funding recommendations to the Texas legislature (THECB, 2013). Other states, such as California,

Virginia and Ohio, provide interactive online tools that can be used to explore median earnings after graduation, by degree level, field, and/or institution (Nye, 2015).

Rigorous evidence regarding the effectiveness of PF policies is limited, but discouraging. Two recent quasi-experimental studies compare trends over time in states adopting new policies to states that did not, and find evidence of unintended strategic responses. Kelchen and Stedrak (2015) find suggestive evidence that colleges under PF may enroll fewer low-income students as a result, while Hillman, Tandberg, and Fryar (2015) find that two-year colleges in Washington increased the production of short-term certificates, but not associate's degrees, when completion rates were introduced as a performance metric. A broader review of the literature by Hillman (2016) identifies 12 studies, which find mostly null or even negative results of PF policies.

As indicated by the failure of many early performance funding programs in the late 1990s, successful design of such programs requires close examination of the mission of institutions, the type of student body it serves and institutional capacity for organizational learning and change (Li, 2014; Dougherty et al., 2014). Alignment with state and social priorities for higher education is crucial, as is confidence in the reliability of the chosen metrics. As more states begin to use labor market data for accountability, it is essential to understand the implications of alternative metrics as well as the potential for unintended consequences, to avoid repeating the mistakes of earlier efforts at reform.

III. Conceptual and practical challenges to using state labor market data

As more and more states are able to track students into the labor market via state UI databases, it opens the door to use this information for institutional accountability. Such use faces a number of important practical and conceptual challenges, however. Practical challenges arise from both mundane data limitations—limited length of follow-up, for example, or an

inability to track graduates out of state—as well as from the fundamental statistical difficulty of disentangling differences in institutions’ true productivity from mere differences in the composition of their respective student populations. Even when stakeholders agree on the outcomes they’d like to measure, these challenges can lead to biased estimates in practice.

Stakeholders may not always agree on what should be measured and when, even if ideal data are available. Conceptual challenges derive from both the multiple objectives that postsecondary institutions serve (for example, improving not just labor market outcomes, but also wellbeing more broadly; promoting degree completion but also access and persistence at other levels) and the multiple purposes and audiences accountability data may be used for (informing enrollment decisions by students, short-term funding decisions by the state, and longer-term strategic planning by institutions). This section describes these challenges, and helps motivate the variety of metrics that we create and compare in the subsequent analysis.

A. Productivity versus student composition

It is one thing to simply measure student outcomes, and another thing entirely to assign all credit (or blame) for those outcomes to the institution. Students at highly selective institutions are likely to have higher graduation rates and better labor market outcomes at least in part because their students come in with stronger academic preparation, better family supports, and greater financial resources. Similarly, student preferences may drive differences in outcomes: students at institutions with strong math and science programs may have better outcomes because math and science majors have better outcomes in general, regardless of the strength of the institution. Finally, students who attend institutions in strong labor markets may have higher earnings than those in weaker labor markets (Hoxby [2015] distinguishes these last two type of selection bias as horizontal selection, while the first represents vertical selection). Failure to

account for selection in a PF system can lead to both biased estimates of true productivity, as well as adverse incentives for institutions to reduce access, as suggested by Kelchen and Stedrak (2015).

Assessing and addressing the selection or “cream-skimming” problem has been a major focus of performance measurement efforts in other sectors (Heckman, Heinrich, & Smith, 2002; Muriel & Smith, 2011; Staiger et al., 2013). While randomized control trials (RCTs) have been used to circumvent selection bias in the evaluation of job training programs, they are less feasible in the context of evaluating schools or hospitals. Still, these concerns have motivated a small but growing literature that uses rigorous quasi-experimental methods to measure institutions’ true causal effects or “value-added.” In higher education, some studies have relied upon on admissions cutoff policies at a limited number of institutions (Hoekstra, 2009), while others have compared students with similar qualifications who were admitted to the same set of selective schools (Dale & Krueger, 2002, 2011). More recently, Hoxby (2015; 2016) uses a vast dataset combining college admissions test scores, enrollment data, and income data from the U.S. Treasury to estimate institutional value-added, relying upon idiosyncracies both in how schools choose between similar students and in how students choose between similar schools to isolate plausibly causal institutional effects. For a detailed review of the selection challenge and related empirical literature in higher education, see Hoxby (2015) and Executive Office of the President (2015).

Unfortunately, there is no guarantee that state policymakers will have access to both the right data and the right natural experiment to undertake these types of rigorous causal analysis. The one state system currently using a “value-added” approach, the Texas technical college system, simply deducts a fixed minimum amount from observed earnings (corresponding to full-

time full-year employment at the minimum wage; see THECB, 2013). Such a strategy is vulnerable to strategic “cream-skimming” behavior, if colleges shift recruitment away from students with the largest potential benefits and towards those with the highest pre-existing earnings potential.

A more generally feasible state strategy would be to compute institutional “fixed effects” that use regression analysis to control for any differences in student outcomes that are attributable to observable student characteristics, such as age, race/ethnicity, gender, location of residence at entry, and declared major. These regression adjustments may be less transparent, and require more choices to be made by the analyst, than simply presenting unadjusted student outcomes. Moreover, differences in *unobserved* student characteristics (such as ability or motivation) are likely to remain even after observed characteristics are taken into account. This may explain why state and federal tools allowing students to browse earnings by institution/program generally provide simple unadjusted means or medians rather than attempting to control for student characteristics.

In the analysis that follows, we present both unadjusted institutional mean outcomes as well as adjusted outcomes using an increasingly rich set of controls. Even with our richest model, however, we do not attempt to interpret the resulting institutional fixed effects as causal. Nor are we able to identify the method that most closely approximates a causal analysis. Our modest goal is to evaluate how much these choices actually matter in practice.

B. Interstate mobility

A major practical challenge in using state UI databases to measure earnings is that such databases typically include information only for individuals who remain in state (though some

states do have data sharing agreements with border states).² Individuals who leave the state are indistinguishable from those who are in state, but simply not working. This complicates the analysis of both employment rates and earnings: without any adjustments, institutions that send many graduates out of state could be seriously disadvantaged on these outcome measures. Moreover, as mobility accumulates over time, this problem worsens the longer the follow-up period. Grogger (2012) discusses this problem in detail in the context of job training program evaluation, and finds that it can seriously compromise the validity of program impact estimates.

In part to minimize this bias, the states that provide information on graduates' employment and earnings often do so within a relatively short period of time post-graduation (e.g. three to six months), and condition earnings metrics upon at least some level of observed employment. For example, Ohio examines in-state retention (a combination of employment and subsequent educational enrollment) in the fourth quarter of the year for spring graduates. Earnings are considered only for those who have earnings above a minimum level approximating full-time employment.

Examining earnings conditional on some approximation of full-time employment has the advantage of avoiding confounds not just from out of state mobility, but also from individual choices regarding labor force participation (e.g. relating to family formation or continued educational investments). On the other hand, such measures will also miss important effects institutions may have on the likelihood of finding and maintaining stable employment.

Our solution to this is to look at graduates in four subsequent quarters in a focal year. If they show up in the data at all, we make the assumption that they are part of the in-state labor

² In addition, UI databases do not include those who are self-employed, some student employees (e.g., work-study students), railroad workers, some agricultural workers, and federal employees. Despite coverage gaps relative to self-reported survey data, prior research has found UI data to provide comparable estimates of program impacts (Kornfeld & Bloom, 1999).

force. We then examine our measures of full-time, full-year employment, social service sector employment, and unemployment claims only for those who appear in the data in that year. For earnings, we further condition on a proxy measure of full-time, full-year employment (described in more detail in the methodology section below).

C. Timing of outcomes measurement

Measures of employment and earnings from relatively early in the life cycle can be not only noisy, but also potentially biased measures of lifetime earnings. As discussed by Heckman, Heinrich, and Smith (2002), “In [the context of human capital programs], the short-run measurements on which performance standards are based will likely be perversely related to long-run benefits” (p. 780). Those with the highest long-term earnings potential may have lower-than-expected earnings if measured soon after graduation, if they continue to invest in additional skills/education both on and off the job. Moreover, those with the highest earnings potential may optimally spend more time after graduation searching for a good job match. Evidence suggests that the optimal time to measure individuals’ earnings is not until their early thirties to mid-forties (Haider & Solon, 2006). Outcomes measured mere months after graduation may reflect mostly noise, or worse—they could be inversely correlated with outcomes over a longer period of time.

From an accountability perspective, however, long time lags also have their own conceptual and practical limitations. To be useful, accountability metrics should reflect institutional performance from a relatively recent period. In addition, the longer the lag between graduation and labor market observation, the more serious the interstate mobility problem becomes. Since the optimal time lag is far from obvious in this context, we measure labor market

outcomes four years after graduation, but also test variations from 1 year to 7 years post-graduation.

D. Measuring outcomes beyond earnings

Even with ideal data on earnings, a fundamental critique that has been leveled against the use of earnings data for post-secondary accountability is that they fail to capture many other positive impacts of education. For example, institutions that send many graduates into teaching or social service jobs will perform worse on earnings-based metrics than those that send many graduates into finance. Even within a given industry, individuals make tradeoffs between wages and other “job amenities” such that wages alone may be a poor summary of overall labor market success. In addition, policymakers (and individuals) may care more about earnings differences at the bottom of the income distribution than in the middle or at the top, but neither average nor median wage metrics will reflect this. Finally, ideally measures of postsecondary accountability would include not just measures of labor market success, but also measures of health and wellbeing.

State UI databases obviously cannot measure all relevant possible institutional effects. Still, even within UI databases it is possible to construct a more diverse range of metrics to capture dimensions beyond earnings. For example, UI data can be used to look at the stability of employment over time (such as whether individuals are employed full-time for the entire year, or how many employers they have had in a given period). Information on industry of employment can also be used to measure employment in “social service” sectors such as teaching or government. Finally, actual unemployment claims can be examined as a measure of job loss (though in practice, UI claims data is typically held separately from quarterly wages and may

require additional data permissions to merge). We describe the specific additional measures that we create in Section IV below.

E. Outcomes for whom? Graduates versus entrants

Most of the state data tools that provide earnings information by institution and program—such as in California, Florida, Virginia, and Ohio—do so for graduates only, rather than looking at outcomes for all students who enter the institution. The conceptual argument for looking only at graduates is twofold: first, institutions may have limited influence over the earnings of students who drop out, and second, given the vast differences in earnings of graduates versus non-graduates, averaging across both groups may be a poor summary of either group’s typical outcomes. On the other hand, examining the earnings only of graduates may seriously distort institutions’ overall productivity if they graduate only a fraction of entrants. The federal College Scorecard is one data source that provides median earnings for all entrants, not just those that graduate.

Our resolution to this tradeoff is to examine labor market outcomes for graduates only, but to examine these metrics alongside graduation metrics that *are* measured for all students. This avoids the problem of interpreting labor market metrics that muddles both margins, while still holding institutions accountable for both.

One limitation of this strategy is that it will not credit institutions that are particularly effective or ineffective at increasing the earnings of non-graduates relative to graduates. This might occur if an institution has a program that is so effective, students leave to take good jobs even before they graduate, or if an institution’s degrees have a particularly high “sheepskin effect” component, such that the payoff to completing 99% of the degree is far less than 99% of the payoff to completing the degree. In general, however, it seems reasonable to assume that

whatever the earnings payoff to graduating from a given institution, the payoff to attending but not graduating may be proportional to the fraction of the degree that was completed (indeed, empirical evidence on the returns to credits from Kane & Rouse [1999] supports this proportional payoff hypothesis).

IV. Empirical methodology

A. Data and Sample

De-identified data were provided by the Ohio Education Research Center (OERC) under a limited-use, restricted data agreement. The OERC assembles data from multiple state agencies, including the Ohio Board of Regents (OBR) and the Ohio Department of Job and Family Services (ODFJS), into a repository known as the Ohio Longitudinal Data Archive (OLDA).³

From the available data, we requested elements from the Ohio Higher Education Information (HEI) system, including students' demographic characteristics, entrance and enrollment records, major choice, and certificate and degree completion from each of the Ohio's higher education institutions (14 universities with 24 regional branch campuses and 23 community colleges, some of which also have multiple campuses). We also requested elements from the unemployment insurance (UI) data, including quarterly earnings and unemployment claims to enable us to examine students' labor market outcomes. While the OLDA data cover more than a decade, for this project we utilize student data from 2000 through 2007 (to enable

³ The following acknowledgement is required to be stated on any materials produced using workforce or higher education data accessed from the OLDA: This workforce solution was funded by a grant awarded to the U.S. Department of Labor's Employment and Training Administration. The solution was created by the Center for Human Resource Research on behalf of the Ohio Department of Job and Family Services and does not necessarily reflect the official position of the U.S. Department of Labor. The Department of Labor makes no guarantees, warranties, or assurances of any kind, express or implied, with respect to such information, including any formation on linked sites and including, but not limited to, accuracy of the information or its completeness, timeliness, usefulness, adequacy, continued availability, or ownership. This solution is copyrighted by the institution that created it. Internal use, by an organization and/or personal use by an individual for non-commercial purposes, is permissible. All other uses require the prior authorization of the copyright owner.

sufficient follow up of entrants/graduates) and labor market data from 2000 through 2012. We describe some additional sample restrictions below, after providing more detail about our methodology.

The data do not include any measure of students' academic ability upon admission (such as SAT/ACT scores, high school grade point average or test scores, or college entrance or placement exam scores), nor do they include financial aid application data or family income information. The data do include information on financial aid receipt for some years; however, for this project we choose to prioritize elements that are available for all analytic cohorts. We may incorporate this information into subsequent sensitivity analyses.

B. Methods and metrics

This section describes the outcome variables we use, the key analysis groups, and the process we employ to estimate the resulting metrics. After computing regression-adjusted “institutional fixed effects” to account for compositional differences across institutions, we standardize group-level means/fixed effects in order to be able to compare metrics that have different natural scales, and then assess the resulting metrics using correlation matrices and graphical analysis.

Outcomes. We construct four labor market accountability metrics based on cohorts of BA/BS graduates for four-year institutions and cohorts of certificate/degree completers and transfer students for two-year institutions (that is, two year students who transferred to a public four-year institution are grouped with those who earned a credential in the same year/institution). We focus on spring graduates only to simplify our analysis, and examine outcomes in the fourth full calendar year post-graduation (so, for a Spring 2002 graduate, this would be calendar year 2006). We test sensitivity to examining these outcomes earlier or later, from one year to 7 years post-graduation.

To avoid contaminating our estimates with out-of-state mobility (those who move out of state are indistinguishable from those in state but not working), we limit all labor market measures to individuals who have at least some in-state earnings during the focal year. We also limit our labor market measures to those who are not enrolled during the focal year.

While our paper is primarily focused on the potential use of labor market outcomes, we also wanted to compare these to academic outcomes which are more commonly used for accountability purposes. We created several measures including degree completion and transfer rates, cumulative credits attempted and completed, and the ratio of credits completed to credits attempted. But because all of these measures were very highly correlated, we chose to focus on degree completion rates (or completion/transfer, for the two-year sector) as a summary academic measure. Additional details on each outcome and its rationale are below:

- 1) Full-time, full-year employment (proxy). This measure is intended to capture the stable employment margin: what percent of graduates are substantially and consistently engaged in the labor market? We do not have any measure of full-time status or hours worked, so we approximate this as employment in all four quarters of the year, with real earnings in each quarter above an amount roughly corresponding to 35 hours per week at minimum wage.⁴ As noted above, this is computed only for individuals who have at least one quarter of positive earnings and are not enrolled in the focal year.
- 2) Annual earnings conditional on full-time, full-year employment. This is intended to capture the intensive employment margin. This is the sum of real quarterly earnings, adjusted to constant 2013 dollars. In practice since we cannot observe hours of work, this measure captures both variation in wages as well as variation in hours. Earnings are top-

⁴ The minimum wage for Ohio in 2013 was 7.85 dollars according to the US Department of Labor. Therefore, the average quarterly minimum wage for full-time employees in 2013 was approximately 4,396 dollars.

coded at the 95th percentile and calculated only for individuals who are employed full-time, full-year and who are not enrolled in the focal year.⁵

- 3) Employment in “social service” sectors. The rationale for this measure is to address the critique that earnings are not the only positive outcome of education. This measure gives credit for potential positive social externalities of public/social sector employment; and could also be a way of acknowledging that some sectors offer benefits and job protections that are not captured by wages alone. Since we only have industry codes in the employment data, this is only a rough proxy: we include those working in educational services (NAICS 661, including private, state and local government schools), and the federal, state and local government excluding state and local schools and hospitals (NAICS 999). For those who have positive earnings (and are not enrolled) at some point within the focal year, we count them as employed in this sector if they worked at least one quarter during that year in one of these selected industries.
- 4) Percent ever claiming unemployment since graduating. This is intended to capture particularly negative employment outcomes that might carry additional weight in policymakers’ social welfare function and might not be captured by average earnings. This is computed only for those who show up in the employment data at some point within the focal year.⁶ UI claims data are only available from 2004 to 2012, therefore this metric has only been estimated for two cohorts of graduates. As opposed to the other

⁵ We considered using median earnings instead of average earnings to diminish the role of outliers. However, medians are more cumbersome to work with in our regression-adjusted models. In sensitivity testing not shown, we found that average earnings after top-coding are very similar to medians, so we stick with averages for simplicity.

⁶ In addition to helping address concerns about out-of-state mobility, this also helps address another concern: individuals cannot claim UI unless they have worked enough in the past year to meet minimum eligibility criteria. This could introduce some ambiguity about whether claiming UI might actually be considered a good outcome, particularly among marginally-attached workers. Our extract of the data do not contain the details necessary to precisely determine UI eligibility; however, in our data about two thirds of those who worked at all during the year have earnings suggesting they are likely to be eligible (in Ohio, individuals must have at least 20 weeks of work in the past year with average weekly earnings of at least \$243).

outcomes, this is a cumulative metric and thus is not restricted by enrollment status within the focal year.

- 5) Degree completion (or transfer) rates. For four-year first-time degree-seeking entrants, we examine BA/BS completion within 6 years of entry. For two-year first-time degree seeking entrants, we include completion of any credential, including short-term certificates (less than one year), long-term certificates (more than one year) and associate's degrees, as well as students who transferred to a four-year institution within three years of entry. We count students as completers regardless of whether they completed at their entry institution. Note, however, that the data only track students in public Ohio institutions, so students who transfer to a private institution or out of state will not be counted here.

Key analysis groups. For the labor market metrics, we use 6 cohorts of baccalaureate and sub-baccalaureate graduates/transfers who earned their first degree or certificate (or transferred, for two-year students) between 2000 and 2005 school years. We examine baccalaureate and sub-baccalaureate institutions separately in all analyses.

For our academic metric, we use 8 cohorts of first-time college students in 2000-2007, admitted as first time undergraduates between age 15 and 60. Students enrolled in a four-year institution whose academic intentions at first entry were to obtain a certificate or AA/AS were excluded from this sample.

Given that in the HEI system baccalaureate degrees awarded are recorded at the institution level and our analysis is at the *campus* level, we use the last campus of enrollment before earning the first BA/BS degree.⁷ We restrict our sample to Ohio residents.⁸ We further

⁷ In the remainder of the analysis, we use institution and campus interchangeably to refer to campus-level estimates, unless specifically noted.

exclude students in the BA/BS sample who were enrolled in a 2-year institution during their last semester of enrolment (this is not many students and simplifies our analysis).

This sample consists of 172,541 baccalaureate students from 39 four-year main and regional branch campuses and 79,255 sub-baccalaureate students from 32 two-year colleges and campuses (which include community colleges, technical colleges and state community colleges). Finally, however, we exclude from our analysis two medical institutions and some small campuses that had fewer than 100 students in the analysis sample for all outcomes. This brings the number of campuses to 30 at the BA/BS level and 28 for the two-year sample.

Computing mean outcomes. The first and simplest thing to do once outcomes are constructed is to compute mean outcomes by campus. It is also straightforward to compute them by program or program-campus; for simplicity we focus on campus. An obvious concern, however, is that differences in outcomes across campus will reflect many factors other than institutional performance: they could reflect differences in students' fields of study, background characteristics (age, race, gender), or differences in local labor markets. This suggests the need to adjust these observed means for compositional differences, a process we describe below.

Computing regression-adjusted institutional fixed effects. The institutional "fixed effect" is simply the estimated contribution of the institution to students' outcomes after accounting for other factors via regression analysis. If no other factors are included in the regression, the fixed effect is equivalent to the unadjusted institutional mean. Our most complete regression model (run separately for two-year and four-year institutions) is the following, run on the individual-level data (we run this without a constant, in order to estimate a full set of institution fixed effects):

⁸ In the event zipcode at entry is missing, we assume individuals are residents as long as they are not otherwise identified as an international student.

$$(1) y_i = instFE + majorFE + \delta X_i + ZLchrs + \gamma_c + \varepsilon_i$$

where i indexes individuals, y_i is a labor market or academic outcome, $instFE$ is a vector of institutional fixed effects (entered as a set of dummy variables indicating the institution initially attended), $majorFE$ is a vector of discipline areas (measured upon college entry) using the two-digit CIP major category,⁹ X_i is a vector of individual background characteristics including gender, race/ethnicity, age, and dummy variables for missing values in student characteristics, and $ZLchrs$ is a vector of 5-digit zip code characteristics taken from the 2007-2011 ACS five-year estimates that include economic and demographic characteristics.¹⁰ Cohort fixed effects, γ_c , are also included to ensure that institutions' graduates (or entrants) are compared against others graduating (or entering) in the same year. We add these covariates in groups to help understand which appear most important, starting with student-level demographics and cohort fixed effects, then adding zip-code level controls. Because college major is not necessarily a fixed student characteristic, but is potentially endogenously influenced by institutions, we add majors last. We note, however, that majors declared at entry are potentially less influenced by institutions than degree fields measured at graduation.

Controlling for ZIP code level characteristics is a way to account both for regional differences in family wealth/SES, which we have no other way to capture, as well as to account

⁹ We use the 2010 Classification of Instructional Programs (CIP) list to create discipline areas. Based on the CIP list, we have the following discipline areas: arts & humanities, business, education, engineering, health, law, natural science & mathematics, services, social & behavioral sciences, and other which includes trades and repair technicians and undeclared/interdisciplinary. Note that we exclude individuals with missing majors at entry, less than 2 percent of the sample.

¹⁰ Five-digit ZIP codes were reported on the admissions application and merged with Census data. These ZIP code characteristics include percent unemployment, percent in labor force, median household income, per capita income, percent of people below poverty line, median age, percent of White, African American and other ethnicities, total population of Hispanics, total population 18 to 24 years old, total population 25 years and older, percent population with less than 9 years of schooling, percent population with 9 to 12 years of schooling, percent population with high school, percent population with some college, percent population with associate's degree, percent population with less than 9 years of schooling, and percent population with less than 9 years of schooling.

for differences in local labor markets.¹¹ Note ZIP codes are measured at initial enrollment, not the time of actual employment. This is preferable because controlling for location at employment (which we do not have in any case) could potentially absorb some of the real impacts of a successful education, if graduates migrate to stronger labor markets in-state.

For first-time college students enrolled in a two-year institution, we also include fixed effects for different categories of students' declared intent at entry (e.g. upgrade skills, train for a new career, transfer before completing, obtain a AA/AS degree). Note that for the academic metrics we are using age at entry, while for labor market metrics we use age as reported at graduation.

Standardizing institutional means/fixed effects. Once the institutional fixed effects are estimated, an entirely separate challenge is what to do with them. It can be particularly difficult to detect patterns across metrics when the metrics are all in different natural scales. While the simplest solution might be to simply rank the institutions on each metric and compare the ranks, this is also limiting because the ranks eliminate valuable information on how far apart institutions are from each other—a small difference in ranks could represent a huge difference in institutional outcomes for some measures but not others, or could represent large difference in the tails of the distribution but not in the middle.

We thus take the middle path of standardizing the institution-level fixed effects by subtracting the overall mean and dividing by the standard deviation. The result is a standardized rating metric that expresses how far above or below the mean the institution is, in standard deviation units for that outcome. This allows us to more easily compare across our different metrics, but note that it produces inherently relative ratings. If policymakers were to use this

¹¹ Alternatively, we could control for ZIP code fixed effects (and we did so in a prior version, with broadly similar results). A potential concern with ZIP code fixed effects, however, is that this may absorb some of the true institutional effects particularly for institutions that attract a predominantly local population.

standardization process in practice, it may make sense to standardize using the mean and standard deviation for an earlier cohort, so that institutions could show improvement over time. Note that this standardization is performed separately for four-year and two-year institutions.

V. Results

A. Baccalaureate institutions

Role of adjustments. To first explore the role of compositional adjustments, Tables 1-4 present, for each of our four labor market metrics, unadjusted institution means side by side with institution fixed effects measured with increasingly rich student-level controls. For ease of comparability across models and outcomes, the institution fixed effects are standardized to mean zero and standard deviation of one. Note that since Model (1) contains no controls, the standardized fixed effects in this column are identical to the standardized raw means. Each table also shows, near the bottom, how each set of metrics correlates with our most fully adjusted model. The pattern of correlations indicates which analytic choices are particularly consequential for the resulting ratings, and which are not.

Several interesting findings emerge from these tables. In general, adjusted and unadjusted metrics are very highly positively correlated. Across our four metrics, the correlation between the unadjusted effects and fully adjusted effects ranges from 0.77 (for our social service employment metric) to 0.97 (for our full-time, full-year employment proxy and for our UI receipt metric). For earnings, the correlation is 0.91. Zipcode-level controls appear the least important controls, across all outcomes: the correlations between Models 2, and 3 are 0.97 or higher across the board. Controlling for field of study makes a particularly large difference for social sector employment.

Even high correlations, however, can mask substantial movement in institutions' ratings and rankings. For our earnings metric, for example, the average institution swung by about ½ of a standard deviation across the four models (i.e., from its most favorable rating to its least favorable rating), or by 5 positions in the rankings of these 30 institutions.¹² One institution's rank swung by 11 positions depending upon which controls were added. Rankings and ratings based on social sector employment rates were similarly volatile. Adjustments matter less for the full-time full-year employment proxy and UI receipt metric: the average swings were only about 0.39 and 0.20 of a standard deviation respectively (or about 3 positions in the rankings in both cases).

Correlations across metrics. Table 5 and Figure 1 examine the relationship between our five different metrics, using estimates from the fully adjusted model. In Figure 1, each vertically-aligned set of points represents an institution's rating on one of our five measures (standardized to mean zero and s.d. of one, to enable comparisons across metrics). If a point lies above zero, that indicates the institution rates above average on that metric. A point at -2, on the other hand, would indicate an institution fell two standard deviations below the institutional mean for that metric. To the extent all points for a given institution are very tightly clustered, that indicates consistency in the institution's rating across metrics. If the points are very far apart vertically (i.e. for a given institution), it means that an institution's rating could be dramatically different depending upon the measure used. To help reveal patterns in the data, the graph is sorted by the degree completion metric, with the lowest ranking institution on this metric on the left and the highest ranking on the right. This makes it easy to identify how top institutions on this metric fare on the labor market metrics, and vice versa.

¹² Even just considering the first three models, which are all correlated at 0.93 or above for the earnings outcome, the average institution swung by 0.32 standard deviations or 3 positions in the rankings of these 30 institutions.

Most metrics are positively correlated with each other, though not often not significantly so; they do seem to capture different information. Degree completion rates correlate positively with all other metrics, but the correlation is only significant for our (reverse coded) UI metric, at 0.33 (the other correlations hover around 0.2). Social sector employment is negatively correlated with both employment and earnings (though not significantly so). Interestingly, the correlation between social sector employment and rates of UI receipt is almost as high as the correlation between employment and earnings (both around 0.49).

In practice, the average difference between an institution's rating on the degree completion metric and its rating on a given labor market metric ranges from 0.93 to 1.01 standard deviations; the average swing across all five metrics is a full 2 standard deviations (or 17 positions in rank). Even just among the four labor market metrics, the average swing is 1.6 standard deviations.

This seemingly large variation in ratings for a single institution is depicted visually in Figure 1. Each institution's ratings (in standard deviation units) are plotted along a vertical line, and the institutions are sorted from left to right by degree completion rates. The graph illustrates both the general positive correlation among the metrics, as well as the dispersion for each institution. The graph also highlights that the dispersion of the labor market metrics is much greater for institutions with low degree completion rates than for those with high degree completion rates.¹³

¹³ We also examined versions of Table 5 and Figure 1 using the raw (unadjusted) versions of our five metrics. Overall, whether or not controls are included does not make much difference to the cross-metric correlations. The exception to this is the social sector employment measure, likely because of its sensitivity to field of study controls. When no controls are included, this measure is significantly negatively correlated with earnings and no longer significantly correlated with our UI metric (though the positive direction remains the same).

Correlations of metrics over time. How sensitive are these labor market metrics to different lengths of follow up? Tables 6 and 7 explore this question from different angles. First, we examine the correlation of the same metric measured different points in time. Table 6 shows that our adjusted measures are generally positively and significantly correlated over time, with the social sector employment metric having the greatest stability and the full-time employment proxy having the least. In the case of the full-time employment proxy: the 1-year and 7-year metrics are barely significantly correlated ($\rho=0.33$), suggesting these measures may be quite misleading if measured too soon after graduation.¹⁴ Again, however, it is important to note that even strongly significant positive correlations can mask important variation: the correlations of the same metric over time are generally much lower than the correlations between adjusted and unadjusted versions of a given metric at a single point in time.

Table 7 and Figure 3 suggest that at least for the conditional earnings metric, analysts may be able to choose between performing statistical adjustments and following graduates for a longer period of time, depending upon which is more feasible. Just between year 1 and year 2, the correlation between the adjusted and unadjusted earnings metric grows from 0.78 to 0.83, and then to 0.91 by year 4. Figure 3 further illustrates this. The figure shows that unadjusted earnings after one-year (the most common and feasible way to measure earnings) are only modestly correlated ($\rho=0.38$) with a fully adjusted metric after 7 years (which might be the most preferred measure except for the inconvenience of waiting that long). However, adjusting the earnings measure after one year is just as good as using an unadjusted measure after two years: both improve the correlation with 7-year earnings to 0.54.

¹⁴ We find broadly similar patterns when we perform the same analysis with the unadjusted metrics.

B. Sub-baccalaureate institutions

Role of adjustments. We repeat the same set of analyses, but this time for sub-baccalaureate institutions. Tables 8-11 present unadjusted and adjusted versions of each of our four labor market metrics, measured 4 years post-graduation. Again, adjusted and unadjusted versions are always positively correlated, ranging from 0.84 (for our earnings metric) to 0.95 (for our social service employment and UI metrics). The typical ratings swing across model adjustments ranges from about 1/3 of a standard deviation for the full-time employment proxy, UI, and social sector employment metrics, to about 1/2 of a standard deviation for the earnings metric. In general, this is quite similar to what we found in the four-year sector.

Correlations across metrics. Table 12 and Figure 3 examine the relationships between our five metrics, using estimates from our fully adjusted model.¹⁵ In notable contrast to the four-year sector, we see here that institution-level earnings and employment are both strongly *negatively* correlated with our completion metric (around -0.53 for both metrics) while our measure of social sector employment is positively but not significantly correlated (0.18).¹⁶ The average ratings swing across these five metrics is 2.2 standard deviations.

This vast difference in institutional ratings dependent on the measure is reflected graphically in Figure 3. As shown, institutions with the highest completion/transfer rates typically have some of the lowest employment and earnings. This striking negative correlation may reflect two issues. First, our completion measure combines certificate completion, AA/AS completion, and transfer to a four-year within three years. If we separate these out (not shown),

¹⁵ We find broadly similar patterns if we use unadjusted versions of these metrics.

¹⁶ Another notable finding is that the UI metric is not as strongly correlated with the other labor market metrics in the two-year sector, as compared with the four-year sector. This may be because of the issue raised earlier, that to claim UI graduates have to have held a job for at least 20 weeks with average earnings above \$243 per week. An institution could thus do “well” on this measure either because its graduates are rarely unemployed, or because they rarely work long enough to qualify for unemployment benefits. We thank Lawrence Katz for raising this point.

long certificate completion (for programs 1-2 years in length) is the most negatively correlated with subsequent employment and earnings, though associate's degree completion is also significantly negatively correlated with earnings. Transfer rates are only slightly positively correlated with earnings four years after transfer (though these correlations are not significant). Second, although our labor market measures exclude students who are currently enrolled in the focal year, these patterns may nonetheless reflect the fact that students who graduate or transfer may still have spent significant time engaged in school in the intervening periods, and thus may have accumulated less work experience, than students who drop out. Yet we find that these negative correlations are still strong if we look 7 years post-graduation (not shown). Overall, the negative correlations between sub-baccalaureate completion rates and subsequent labor market outcomes is puzzling and provides strong motivation for considering measures beyond graduation/transfer for this sector.

Correlation of metrics over time. Table 13 examines the sensitivity of these metrics to the length of follow up. It appears that labor market outcomes are much less sensitive to length of follow up for the two-year sector than we found for the four-year sector.¹⁷ Table 14 further shows that unlike with the four-year sector, it is not obvious that the role of statistical adjustments diminishes over time for this sector. The correlation between adjusted and unadjusted versions of the same metric appears more stable, and sometimes even declining over time.

VI. Discussion

While newly accessible state UI databases present great opportunities for enhancing states' ongoing efforts to measure college student outcomes, it is no straightforward task to

¹⁷ This holds regardless of whether we use adjusted or unadjusted measures.

figure out how to use these data most effectively. We draw the following conclusions from our analyses:

First, state UI databases can provide richer measures of graduates' labor market experiences beyond just earnings. While three of the four labor market metrics are positively correlated (with social sector employment the exception), they do appear to capture different aspects of post-college labor market success, and institutions could receive markedly different ratings or rankings depending upon which measure is used.

Second, metrics based upon labor market outcomes result in substantially different institutional ratings and rankings than those based on degree completion (or completion/transfer) alone, particularly in the two-year sector. Indeed, for two-year institutions, degree completion/transfer rates are negatively correlated with 3 of our 4 labor market outcome metrics, highlighting the risk involved in relying on academic outcomes alone.

Third, statistical adjustments generally have less consequence for ratings/rankings than choice of outcome metric and length of follow up. Moreover, field of study controls appear particularly important. Research in the Canadian context by Betts, Ferrall, and Finnie (2013) similarly highlights the role of field of study controls, and it is worth noting that data on major field of study are one of the comparative advantages of state administrative databases compared to alternative national data sources (such as IRS data). Still, overall the effect of adjustments appears modest compared to other analytic choices. It is possible that even our fully adjusted model omits important factors that if incorporated, could more substantially change institutions' ratings. But our finding echoes a similar pattern in hospital performance measurement, in which the choice of outcome generally matters more than which patient-level controls are included (Staiger, 2016).

Fourth, for earnings-based metrics in the four-year sector, statistical adjustments appear more important when outcomes are measured early. Compared against 7th-year adjusted earnings, an adjusted one-year measure performed about as well as an unadjusted two-year measure. This suggests states may be able to choose between using an adjusted measure soon after graduation, or an unadjusted measure after a longer period of time, depending upon which is more feasible. This tradeoff is not evident for every outcome metric, however, or for the two-year sector.

Finally, when we examine the correlation of our metrics over different lengths of follow up, we find that our conditional earnings metric is much less stable over time for the four-year sector than for the two-year sector. In the four-year sector the correlation of 7 year earnings with earnings measured earlier ranges from 0.55 to 0.87 while in the two-year sector the equivalent correlations range from 0.88 to 0.93. The full-time, full-year employment metric is even more unstable for four-year graduates: the correlation between employment measured at 1 year versus 7 years post-graduation is only 0.331, and only marginally significant.

Limitations. Currently we use several cohorts of entrants/graduates to estimate each institution's fixed effect. We have not yet examined what would happen if these effects were estimated with only one or two cohorts at a time. We have not incorporated any controls for student ability or family income, which have been used in other studies of accountability metrics. Finally, we have not examined any input-based measures of institutional quality/selectivity, such as constructed in Dillon and Smith (2015). It would be very valuable to further investigate the correlation between input- and output-based institutional ratings, as done by Betts, Ferrall, and Finnie (2013) in the Canadian context.

Conclusion. Overall, our preliminary conclusion is that labor market data, even when imperfect, can provide valuable information distinct from students' academic outcomes, particularly for the two-year sector. Institutional ratings based on labor market outcomes, however, are quite sensitive to the specific metric constructed. The simplest labor market metrics at policymakers' disposal— unadjusted employment rates and average earnings within a year after graduation — both prove to be quite unreliable compared to the same outcomes measured later. Moreover, earnings and employment on their own may fail to capture other aspects of economic wellbeing of value to both policymakers and students themselves. Consistent with similar types of studies conducted in other contexts (such as outcomes-based evaluations of hospital quality), the choice of metric and length of follow up appear to matter more than compositional adjustments. Overall, our findings suggest a cautious approach: while a mix of feasible labor market metrics may be better than none, reliance on any one metric—particularly one measured early—may unintentionally undermine policymakers' ongoing efforts to accurately quantify institutional performance.

References

- Barnow, Burt S., and Jeffrey Smith. Employment and Training Programs. No. w21659. National Bureau of Economic Research, 2015.
- Betts, Julian, Christopher Ferrall, and Ross Finnie. (2013). The Role of University Characteristics in Determining Post-Graduation Outcomes : Panel Evidence from Three Canadian Cohorts. *Canadian Public Policy*, 39 : S81-S106.
- Bogue, E. G., & Johnson, B. D. (2010). Performance incentives and public college accountability in the United States: A quarter century policy audit. *Higher Education Management and Policy*, 22(2), 9–30.
- Burke, J. C. (2001). Accountability, reporting, and performance: Why haven't they made more difference? New York: *Ford Foundation*.
- Burke, J. C. (Ed.). (2002). Funding public colleges and universities for performance: Popularity, problems, and prospects. Albany, NY: *Rockefeller Institute Press*.
- Burke, J. C., & Serban, A. M. (1997). Performance funding and budgeting for public higher education: Current status and future prospects. Albany, NY: *Rockefeller Institute*.
- Burke, J., & Modarresi, S. (2001). Performance funding programs: Assessing their stability. *Research in Higher Education*, 42(1), 51–71. Retrieved from <http://link.springer.com/article/10.1023/A:1018764511093>
- 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.
- Dillon, Eleanor W. & Jeffery A. Smith (2015). *Determinants of the Match between Student Ability and College Quality*. Working paper, Arizona State University. URL: http://www.public.asu.edu/~edillon2/Dillon_Smith_Determinants.pdf
- Dougherty, K. J., & Reddy, V. (2011). The impacts of state performance funding systems on higher education institutions: Research literature review and policy recommendations. New York: *Community College Research Center*, Teachers College, Columbia University. Retrieved from

<http://ccrc.tc.columbia.edu/publications/impacts-state-performance-funding.html>

- Dougherty, K. J., & Reddy, V. (2013). Performance funding for higher education: What are the mechanisms? What are the impacts? (ASHE Higher Education Report). San Francisco, CA: Jossey-Bass.
- Dougherty, K. J., Hare, R. J., & Natow, R. S. (2009). Performance accountability systems for community colleges: Lessons for the Voluntary Framework of Accountability for Community Colleges. New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from <http://ccrc.tc.columbia.edu/publications/performance-accountability-systems.html>
- Dougherty, K. J., Natow, R., & Vega, B. (2012). Popular but Unstable: Explaining Why State Performance Funding Systems in the United States Often Do Not Persist. *Teachers College Record*, 114(030301), 1–41.
- Dougherty, Kevin J. & Rebecca Natow. (2010). Change in Long-Lasting State Performance Funding Systems for Higher Education: The Cases of Tennessee and Florida. Working Paper #18. New York: *Community College Research Center*, Teachers College, Columbia University.
- Dougherty, Kevin J., Sosanya S. Jones, Hana Lahr, Rebecca S. Natow, Lara Pheatt, and Vikash Reddy. (2014) Implementing Performance Funding in Three Leading States: Instruments, Outcomes, Obstacles, and Unintended Impacts. New York, NY: *Community College Research Center Working Paper No. 74*, 2014.
- Dougherty, Kevin J., Sosanya S. Jones, Hana Lahr, Rebecca S. Natow, Lara Pheatt, and Vikash Reddy. (2014). The Political Origins of Performance Funding 2.0 in Indiana, Ohio, and Tennessee: Theoretical Perspectives and Comparisons with Performance Funding 1.0. New York, NY: *Community College Research Center Working Paper No. 68*, 2014.
- Executive Office of the President (EOP). (2015). Using Federal Data to Measure and Improve the Performance of U.S. Institutions of Higher Education. *The Executive Office of the President*. Retrieved from: <https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAndImprovePerformance.pdf>
- Florida State University System Board of Governors [FLBOG]. (2016). *Board of Governors Performance Funding Model Overview*. Tallahassee: FLBOG. URL: http://www.flbog.edu/about/budget/docs/performance_funding/Overview-Doc-Performance-Funding-10-Metric-Model-Condensed-Version.pdf. Accessed May 26, 2016.
- Grogger, Jeffrey. (2012). Bounding the Effects of Social Experiments: Accounting for Attrition in Administrative Data. *Evaluation Review*, 36(6): 449-474.

- Haider, Steven and Gary Solon. (2006). Life-Cycle Variation in the Association between Current and Lifetime Earnings. *American Economic Review*, 96(4): 1308-1320.
- Hastings, Justine S., Christopher A. Neilson, Anely Ramirez, and Seth D. Zimmerman. (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.
- Heckman, James J., Carolyn Heinrich and Jeffrey Smith. (2002). The Performance of Performance Standards. *Journal of Human Resources*, vol. 37(4):778-811.
- Hillman, Nicholas. (2016). *Why Performance-Based College Funding Doesn't Work*. New York: The Century Foundation.
- Hillman Nicholas W., David A. Tandberg and Alisa H. Fryar. (2015). Evaluating the Impacts of “New” Performance Funding in Higher Education. *Educational Evaluation and Policy Analysis* 37(4):501–519.
- 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.
- Hoxby, Caroline M. (2015). Computing the Value-Added of American Postsecondary Institutions. Internal Revenue Service. Retrieved from <https://www.irs.gov/pub/irs-soi/15rpcompvalueaddpostsecondary.pdf>
- Jacobson, Louis and Robert LaLonde. (2013). Using Data to Improve the Performance of Workforce Training. The Hamilton Project and Results for America.
- 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.
- Kelchen, R., & Stedrak, L. J. (2016). Does Performance-Based Funding Affect Colleges’ Financial Priorities? *Journal of Education Finance*, 41(3), 302-321.
- Layzell, D. T. (1999). Linking performance to funding outcomes at the state level for public institutions of higher education: Past, present and future. *Research in Higher Education*, 40(2), 233-246.
- Kornfeld, Robert and Howard S. Bloom (1999). Measuring Program Impacts on Earnings and Employment: Do Unemployment Insurance Wage Reports from Employers Agree with Surveys of Individuals? *Journal of Labor Economics* 17(1): 168-197
- Li, A.Y. (2014). Performance Funding in the States: An Increasingly Ubiquitous Public

- Policy for Higher Education. *Higher Education in Review*, 11.
- National Conference of State Legislatures. (2015). Performance funding for higher education. Denver, CO: Author. Retrieved from <http://www.ncsl.org/issues-research/educ/performance-funding.asp>
- Ohio Association of Community Colleges (2013). SSI allocation recommendations. Columbus, OH: Author.
- Nye, J., Isaac Rowlett, Gordon Sonnenschein, Anika Van Eaton, and Sarah Wissel (2015). Connecting Community College Funding to Workforce Outcomes: An Assessment of the National Landscape. The George Washington University.
- Muriel, Alastair and Jeffrey Smith. (2011). On Educational Performance Measures. *Fiscal Studies*, 32 (2): 187-206.
- Snyder, Martha. (2011). Performance Funding in Indiana: An Analysis of Lessons from the Research and Other State Models. Washington, DC: *HCM Strategists*, 2011.
- Snyder, Martha. (2015). Driving Better Outcomes: Typology and Principles to Inform Outcomes-Based Funding Models. Washington, DC: *HCM Strategists*, 2015.
- Selingo, Jeffrey & Martin Van Der Werf. (2016). *Linking appropriations for the Texas State Technical College System to Student Employment Outcomes*. Indianapolis, IN: Lumina Foundation.
- Staiger, Douglas O., Lena Chen, John Birkmeyer, Andrew Ryan, Wenying Zhang, and Justin Dimick. (2013). Composite Quality Measures for Common Inpatient Medical Conditions. *Medical Care*, 2013, 51(9):832-837.
- Staiger, Douglas O. (2016). *What Healthcare Teaches Us about Measuring Productivity in Higher Education*. Presentation given at NBER Productivity in Higher Education Conference, Cambridge, MA, June 1, 2016.
- SRI International. (2012). States' Methods of Funding Higher Education. Menlo Park, CA.
- Texas Higher Education Coordinating Board (2013). *The Texas State Technical College System Returned Value Funding Model Methodology*. Austin, TX: Texas Higher Education Coordinating Board (July 2013).
- Umbricht, Mark R., Frank Fernandez, and Justin C. Ortagus. (2015). An Examination of the (Un)Intended Consequences of Performance Funding in Higher Education. *Educational Policy*: 1—31. doi: 10.1177/0895904815614398.

U.S. Department of Education. (2013). Education Department Releases College Scorecard to Help Students Choose Best College for Them. Retrieved from <http://www.ed.gov/news/press-releases/education-department-releases-college-scorecard-help-students-choose-best-college-them>

Wiswall, Matthew and Basit Zafar. (2015). How Do College Students Respond to Public Information about Earnings? *Journal of Human Capital* 9(2): 117-169.

Table 1. Institutional Fixed Effects: Conditional Earnings, Year 4 Post-Graduation*Baccalaureate institutions*

Institution Code	Raw Mean	<u>Model (1)</u>	<u>Model (2)</u>	<u>Model (3)</u>	<u>Model (4)</u>
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_17	\$52,988	1.03	1.50	1.44	1.97
camp_06	\$55,183	1.68	2.16	2.19	1.44
camp_04	\$55,099	1.66	1.83	1.89	1.38
camp_12	\$52,755	0.96	0.54	0.64	1.34
camp_16	\$51,642	0.63	0.84	0.84	1.20
camp_14	\$53,747	1.26	0.24	0.27	0.90
camp_27	\$53,547	1.20	1.39	1.45	0.87
camp_19	\$52,127	0.78	1.05	0.99	0.73
camp_07	\$51,747	0.66	0.73	0.67	0.55
camp_18	\$50,318	0.24	0.32	0.25	0.52
camp_23	\$49,489	-0.01	0.26	0.18	0.48
camp_20	\$49,787	0.08	-0.22	-0.09	0.44
camp_15	\$51,335	0.54	-0.45	-0.39	0.11
camp_01	\$51,061	0.46	0.53	0.47	0.08
camp_29	\$49,699	0.05	0.29	0.30	-0.08
camp_03	\$50,384	0.26	0.31	0.19	-0.11
camp_10	\$49,126	-0.12	-0.25	-0.25	-0.27
camp_22	\$46,913	-0.77	-0.62	-0.65	-0.34
camp_24	\$44,895	-1.37	-1.34	-1.09	-0.39
camp_02	\$47,383	-0.63	-0.35	-0.52	-0.44
camp_28	\$47,174	-0.70	-0.60	-0.73	-0.45
camp_05	\$49,853	0.10	-0.19	-0.31	-0.53
camp_25	\$48,159	-0.40	-0.57	-0.47	-0.60
camp_11	\$47,424	-0.62	-0.34	-0.33	-0.63
camp_26	\$47,899	-0.48	-0.58	-0.26	-0.83
camp_30	\$48,056	-0.43	-0.52	-0.51	-0.88
camp_09	\$48,276	-0.37	-0.96	-1.00	-0.97
camp_21	\$44,113	-1.60	-1.39	-1.51	-1.32
camp_13	\$45,029	-1.33	-1.62	-1.67	-1.80
camp_08	\$40,293	-2.74	-1.98	-1.98	-2.37
Institution-level mean	\$49,517	0.00	0.00	0.00	0.00
Institution-level std. dev.	\$3,367	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.92	1.00		
Model 3		0.93	0.99	1.00	
Model 4		0.91	0.90	0.91	1.00
Observations	66,695	66,695	66,695	66,695	66,695

Notes: Institutions sorted by Model 4 effects. Earnings are measured for non-enrolled graduates in four consecutive quarters in the fourth full year post graduation, and are measured conditional on our proxy of full-time, full-year employment (see text for additional details), so the overall average of \$49,517 is among those employed full-time, full-year in-state, and not still enrolled in that year.

Table 2. Institutional FE: Full-time Full-year Employment (Proxy), Year 4 Post-Graduation
Baccalaureate institutions

Institution Code	Raw Mean	Model (1)	Model (2)	Model (3)	Model (4)
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_12	0.80	1.97	2.01	2.28	2.48
camp_10	0.77	1.25	1.30	1.67	1.47
camp_20	0.76	0.94	1.21	0.95	1.47
camp_18	0.77	1.38	1.41	1.21	1.21
camp_17	0.75	0.76	0.61	0.63	0.73
camp_07	0.73	0.35	0.56	0.63	0.64
camp_29	0.75	0.71	0.76	0.78	0.64
camp_02	0.75	0.74	0.61	0.52	0.53
camp_11	0.73	0.25	0.13	0.52	0.50
camp_28	0.77	1.28	1.38	0.78	0.47
camp_01	0.74	0.61	0.61	0.58	0.47
camp_06	0.73	0.30	0.61	0.66	0.44
camp_15	0.74	0.53	0.18	0.52	0.32
camp_23	0.73	0.17	-0.04	-0.03	0.26
camp_22	0.71	-0.11	0.01	-0.17	0.11
camp_04	0.73	0.40	0.36	0.37	0.11
camp_16	0.72	0.12	-0.04	-0.09	-0.01
camp_19	0.72	-0.09	-0.24	-0.17	-0.09
camp_27	0.73	0.33	0.27	0.29	-0.09
camp_21	0.70	-0.40	-0.39	-0.84	-0.36
camp_25	0.70	-0.52	-0.47	-0.72	-0.45
camp_08	0.64	-2.10	-0.73	-0.61	-0.89
camp_26	0.70	-0.45	-0.61	-0.64	-0.89
camp_13	0.67	-1.17	-1.44	-1.07	-0.92
camp_05	0.69	-0.63	-0.79	-0.92	-1.04
camp_14	0.68	-1.04	-1.18	-0.87	-1.07
camp_24	0.66	-1.40	-1.44	-1.59	-1.25
camp_30	0.68	-0.91	-1.13	-1.15	-1.31
camp_03	0.66	-1.40	-1.36	-1.67	-1.66
camp_09	0.65	-1.89	-2.15	-1.85	-1.81
Institution-level mean	0.72	0.00	0.00	0.00	0.00
Institution-level std. dev.	0.04	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.95	1.00		
Model 3		0.93	0.97	1.00	
Model 4		0.93	0.96	0.97	1.00
Observations	91,600	91,600	91,600	91,600	91,600

Notes: Full-time, full-year employment is estimated by examining four consecutive quarters in the fourth full year post graduation, and requires a graduate to earn above a minimum amount in each quarter corresponding to 35 hours per week at minimum wage. The sample is restricted to graduates who are not still enrolled have positive earnings in at least one quarter of the focal year.

Table 3. Institutional Fixed Effects: Social Sector Employment (Proxy), Year 4 Post-Graduation
Baccalaureate institutions

Institution Code	Raw Mean	<u>Model (1)</u>	<u>Model (2)</u>	<u>Model (3)</u>	<u>Model (4)</u>
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_20	0.44	1.62	1.60	1.56	1.55
camp_25	0.46	1.82	1.70	1.63	1.47
camp_05	0.35	0.76	0.82	1.05	1.29
camp_24	0.49	2.09	2.07	1.88	1.13
camp_26	0.35	0.77	0.81	0.51	0.97
camp_08	0.36	0.81	0.82	0.90	0.86
camp_22	0.33	0.53	0.43	0.39	0.74
camp_19	0.20	-0.60	-0.40	-0.31	0.45
camp_17	0.27	-0.01	0.22	0.38	0.42
camp_23	0.25	-0.16	0.08	0.13	0.42
camp_21	0.35	0.78	0.63	0.41	0.40
camp_29	0.26	-0.06	0.03	0.04	0.39
camp_07	0.24	-0.28	-0.24	-0.10	0.30
camp_28	0.47	1.91	1.77	1.83	0.06
camp_18	0.35	0.72	0.73	0.85	0.05
camp_01	0.23	-0.38	-0.33	-0.31	0.05
camp_04	0.17	-0.97	-0.84	-0.83	-0.01
camp_02	0.26	-0.07	0.11	0.22	-0.02
camp_11	0.21	-0.51	-0.44	-0.33	-0.07
camp_30	0.25	-0.16	-0.13	-0.24	-0.14
camp_27	0.18	-0.81	-0.65	-0.69	-0.19
camp_06	0.17	-0.94	-1.19	-1.18	-0.32
camp_03	0.29	0.19	-0.08	-0.11	-0.32
camp_13	0.19	-0.75	-0.74	-0.70	-0.43
camp_16	0.20	-0.67	-0.53	-0.53	-0.47
camp_09	0.23	-0.40	-0.61	-0.57	-0.75
camp_10	0.13	-1.28	-1.28	-1.26	-1.06
camp_12	0.11	-1.48	-1.66	-1.80	-1.91
camp_15	0.18	-0.81	-0.76	-0.77	-1.93
camp_14	0.09	-1.65	-1.93	-2.04	-2.93
Institution-level mean	0.27	0.00	0.00	0.00	0.00
Institution-level std. dev.	0.11	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.99	1.00		
Model 3		0.98	0.99	1.00	
Model 4		0.77	0.82	0.83	1.00
Observations	91,600	91,600	91,600	91,600	91,600

Notes: Social sectoremployment is estimated by examining four consecutive quarters in the fourth full year post graduation, and requires a graduate to have been employed in educational services, or government in at least one of these quarters. Sample is limited to graduates who are not still enrolled and who have positive earnings in at least one quarter of the focal year.

Table 4. Institutional Fixed Effects: Cumulative UI Receipt, Year 4 Post-Graduation*Baccalaureate institutions*

Institution Code	Raw Mean	Model (1)	Model (2)	Model (3)	Model (4)
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_17	0.06	1.03	1.01	1.04	1.09
camp_19	0.08	0.70	0.81	0.83	0.91
camp_23	0.07	0.74	0.72	0.77	0.84
camp_22	0.08	0.57	0.58	0.54	0.68
camp_20	0.09	0.49	0.60	0.65	0.68
camp_26	0.09	0.49	0.56	0.74	0.68
camp_04	0.09	0.42	0.51	0.58	0.66
camp_21	0.08	0.68	0.69	0.56	0.57
camp_28	0.06	1.03	0.96	0.72	0.54
camp_29	0.09	0.44	0.45	0.56	0.54
camp_18	0.08	0.64	0.58	0.58	0.52
camp_13	0.10	0.33	0.36	0.29	0.45
camp_12	0.09	0.42	0.51	0.40	0.41
camp_16	0.10	0.21	0.18	0.20	0.27
camp_24	0.10	0.25	0.22	0.27	0.27
camp_07	0.12	-0.08	0.09	0.20	0.25
camp_01	0.10	0.21	0.22	0.22	0.22
camp_02	0.10	0.33	0.20	0.22	0.13
camp_05	0.10	0.23	0.13	0.18	0.11
camp_09	0.12	-0.05	0.13	0.09	0.09
camp_11	0.10	0.18	0.07	0.02	0.02
camp_27	0.11	0.01	0.00	0.09	0.02
camp_30	0.12	-0.03	0.00	-0.03	-0.12
camp_25	0.11	-0.01	-0.02	-0.05	-0.12
camp_06	0.13	-0.34	-0.49	-0.44	-0.41
camp_15	0.16	-0.87	-0.78	-0.89	-0.89
camp_10	0.16	-0.81	-1.01	-1.09	-1.05
camp_03	0.16	-0.77	-0.92	-1.00	-1.26
camp_08	0.30	-3.49	-3.07	-2.81	-2.67
camp_14	0.27	-2.97	-3.29	-3.45	-3.44
Institution-level mean	0.11	0.00	0.00	0.00	0.00
Institution-level std. dev.	0.05	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.99	1.00		
Model 3		0.98	0.99	1.00	
Model 4		0.97	0.99	1.00	1.00
Observations	35,317	35,317	35,317	35,317	35,317

Notes: Institutional fixed effects from Models 1-4 are reverse-coded so that lower rates of UI receipt correspond to more positive standardized FE. Cumulative UI receipt is measured as the percent ever receiving UI, or other unemployment compensation, by the end of the fourth full year post graduation. To limit bias from out-of-state mobility, sample is limited to those with positive earnings in at least one quarter of the focal year.

Table 5. Correlations of Adjusted Institution-Level Metrics*BACCALAUREATE INSTITUTIONS*

Correlations	ba6yr	Ftemp 4yrs	Earn 4yrs	SS sec 4yrs	UI 4yrs
ba6yr	1.000				
Ftemp 4yrs	0.176	1.000			
Earn 4yrs	0.215	0.497***	1.000		
SS sec 4yrs	0.226	-0.117	-0.277	1.000	
UI 4yrs	0.325*	0.316*	0.202	0.492***	1.000
Avg diff. vs. BA metric (sd's)	0.00	1.01	1.00	0.97	0.93
Avg diff. vs. earnings metric (sd's)	1.00	0.86	0.00	1.24	0.94

Note: ***=p<.01, **=p<.05, *=p<.10.

Table 6. Correlations of Adjusted LM Metrics Over Time**A. Full-time employment proxy**

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.713***	1.000		
Year 4	0.592***	0.730***	1.000	
Year 7	0.331*	0.266	0.337*	1.000

B. Conditional earnings

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.933***	1.000		
Year 4	0.765***	0.881***	1.000	
Year 7	0.553***	0.672***	0.874***	1.000

C. Social sector employment

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.982***	1.000		
Year 4	0.965***	0.955***	1.000	
Year 7	0.957***	0.948***	0.963***	1.000

D. UI Receipt (inverse)

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.828***	1.000		
Year 4	0.864***	0.886***	1.000	
Year 7	0.550***	0.745***	0.769***	1.000

Note: ***=p<.01, **=p<.05, *=p<.10.

Table 7. Correlations Between Adjusted and Unadjusted Metrics

Metric (Standardized)	1 Year	Year 2	Year 4	Year 7
Ftemp	0.920	0.956	0.931	0.941
Earn	0.778	0.828	0.905	0.905
SS sec	0.679	0.701	0.770	0.693
UI	0.989	0.978	0.967	0.965

Note: $p < .01$ for all correlations in this table.

Table 8. Institutional Fixed Effects: Conditional Earnings, Year 4 Post-Graduation*Subbaccalaureate institutions*

Institution Code	Raw Mean	Model (1)	Model (2)	Model (3)	Model (4)
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_23	\$46,940	0.37	0.65	0.92	1.65
camp_08	\$49,261	1.07	1.02	0.94	1.27
camp_06	\$49,592	1.17	1.21	1.08	1.26
camp_07	\$49,985	1.29	1.48	1.52	1.00
camp_21	\$46,722	0.30	0.61	0.94	0.98
camp_13	\$49,200	1.05	0.86	0.69	0.97
camp_03	\$48,919	0.97	0.98	0.89	0.81
camp_20	\$48,518	0.85	0.80	0.85	0.71
camp_22	\$48,642	0.88	0.78	0.76	0.64
camp_02	\$47,651	0.58	0.55	0.62	0.52
camp_04	\$47,800	0.63	0.93	0.92	0.44
camp_12	\$47,827	0.64	0.85	0.77	0.42
camp_05	\$47,354	0.49	0.43	0.20	0.22
camp_19	\$47,086	0.41	0.22	0.17	0.12
camp_24	\$42,631	-0.94	-0.90	-0.86	0.02
camp_26	\$47,103	0.42	0.26	0.27	-0.03
camp_16	\$46,797	0.32	0.27	0.24	-0.04
camp_14	\$45,953	0.07	0.24	0.32	-0.13
camp_09	\$44,329	-0.42	-0.65	-0.73	-0.16
camp_27	\$45,173	-0.17	-0.81	-0.88	-0.42
camp_25	\$38,307	-2.25	-1.81	-1.74	-0.54
camp_17	\$45,554	-0.05	-0.18	-0.13	-0.68
camp_18	\$42,365	-1.02	-0.97	-1.00	-0.81
camp_10	\$43,665	-0.63	-0.76	-0.70	-1.08
camp_15	\$40,392	-1.62	-1.65	-1.65	-1.49
camp_28	\$40,049	-1.72	-1.58	-1.59	-1.63
camp_01	\$42,902	-0.86	-1.04	-1.00	-1.77
camp_11	\$39,743	-1.82	-1.79	-1.84	-2.25
Institution-level mean	\$45,731	0.00	0.00	0.00	0.00
Institution-level std. dev.	\$3,297	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.98	1.00		
Model 3		0.96	0.99	1.00	
Model 4		0.84	0.88	0.89	1.00
Observations	36,596	36,596	36,596	36,596	36,596

Notes: Institutional fixed effects from Models 1-4 are standardized to mean 0 and s.d. 1. Institutions sorted by Model 4 effects. Earnings are measured for four consecutive quarters in the fourth full year post graduation, and are measured conditional on our proxy of full-time, full-year employment (see text for additional details), so the overall average of \$45,731 is among those employed full-time, full-year in-state in that year. Sample also restricted to those were not enrolled within the focal year.

Table 9. Institutional FE: Full-time Full-year Employment (Proxy), Year 4 Post-Graduation
Subbaccalaureate institutions

Institution Code	Raw Mean	Model (1)	Model (2)	Model (3)	Model (4)
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_01	0.53	-2.22	-2.29	-2.36	-2.85
camp_02	0.69	0.26	0.27	0.31	0.27
camp_03	0.72	0.76	0.85	0.96	0.74
camp_04	0.71	0.66	0.69	0.54	0.19
camp_05	0.68	0.13	0.19	0.28	0.20
camp_06	0.67	-0.11	0.33	0.71	0.82
camp_07	0.69	0.21	0.63	0.97	0.71
camp_08	0.70	0.42	0.35	0.53	0.68
camp_09	0.72	0.71	0.55	0.25	0.61
camp_10	0.59	-1.26	-1.41	-1.31	-1.57
camp_11	0.50	-2.77	-2.73	-2.69	-2.70
camp_12	0.71	0.61	0.69	0.71	0.61
camp_13	0.68	0.11	0.00	0.13	0.30
camp_14	0.71	0.58	0.58	0.31	0.02
camp_15	0.65	-0.31	-0.37	-0.45	-0.41
camp_16	0.70	0.40	0.35	0.18	-0.14
camp_17	0.72	0.71	0.62	0.51	0.14
camp_18	0.70	0.39	0.35	0.11	0.25
camp_19	0.74	1.06	0.96	0.65	0.51
camp_20	0.72	0.69	0.66	0.73	0.62
camp_21	0.70	0.39	0.44	0.54	0.61
camp_22	0.73	0.97	0.91	1.09	0.96
camp_23	0.66	-0.23	-0.22	-0.19	0.33
camp_24	0.64	-0.53	-0.56	-0.62	-0.01
camp_25	0.60	-1.10	-0.99	-0.86	-0.03
camp_26	0.74	1.13	1.02	1.02	0.78
camp_27	0.69	0.31	0.03	-0.18	0.11
camp_28	0.55	-1.93	-1.92	-1.87	-1.73
Institution-level mean	0.67	0.00	0.00	0.00	0.00
Institution-level std. dev.	0.06	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.99	1.00		
Model 3		0.96	0.98	1.00	
Model 4		0.90	0.93	0.95	1.00
Observations	53,353	53,353	53,353	53,353	53,353

Notes: Institutional fixed effects from Models 1-4 are standardized to mean 0 and s.d. 1. Institutions sorted by Model 4 effects. Full-time, full-year employment is estimated by examining four consecutive quarters in the fourth full year post graduation, and requires a graduate to earn above a minimum amount in each quarter corresponding to 35 hours per week at minimum wage. The sample is restricted to graduates who are not still enrolled and have positive earnings in at least one quarter of the focal year.

Table 10. Institutional Fixed Effects: Social Sector Employment (Proxy), Year 4 Post-Graduation
Subbaccalaureate institutions

Institution Code	Raw Mean	Model (1)	Model (2)	Model (3)	Model (4)
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_01	0.15	0.58	0.59	0.59	0.99
camp_02	0.14	0.47	0.41	0.31	0.18
camp_03	0.11	-0.29	-0.40	-0.51	-0.07
camp_04	0.11	-0.29	-0.17	-0.08	0.31
camp_05	0.18	1.55	1.39	1.34	1.40
camp_06	0.15	0.63	0.04	-0.28	-0.18
camp_07	0.10	-0.64	-1.10	-1.23	-0.89
camp_08	0.11	-0.34	-0.38	-0.31	-0.45
camp_09	0.08	-1.07	-1.03	-0.98	-1.14
camp_10	0.17	1.26	1.26	1.26	0.94
camp_11	0.11	-0.42	-0.27	-0.28	-0.56
camp_12	0.12	-0.04	-0.01	-0.11	-0.02
camp_13	0.11	-0.26	-0.30	-0.23	-0.40
camp_14	0.09	-0.80	-0.66	-0.68	-0.43
camp_15	0.10	-0.72	-0.61	-0.53	-0.92
camp_16	0.12	0.01	0.01	0.11	0.37
camp_17	0.13	0.09	0.07	0.09	0.31
camp_18	0.08	-1.32	-1.13	-1.13	-1.41
camp_19	0.09	-0.91	-0.92	-0.83	-0.75
camp_20	0.11	-0.40	-0.38	-0.46	-0.45
camp_21	0.19	1.72	1.71	1.84	2.09
camp_22	0.13	0.06	-0.04	-0.13	0.07
camp_23	0.11	-0.40	-0.22	-0.13	-0.59
camp_24	0.24	3.02	3.09	3.01	2.80
camp_25	0.14	0.39	0.77	0.91	0.26
camp_26	0.07	-1.51	-1.49	-1.38	-0.86
camp_27	0.09	-1.02	-1.05	-1.06	-1.46
camp_28	0.15	0.63	0.82	0.91	0.86
Institution-level mean	0.12	0.00	0.00	0.00	0.00
Institution-level std. dev.	0.04	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.98	1.00		
Model 3		0.96	0.99	1.00	
Model 4		0.95	0.95	0.95	1.00
Observations	53,353	53,353	53,353	53,353	53,353

Notes: Institutional fixed effects from Models 1-4 are standardized to mean 0 and s.d. 1. Institutions sorted by Model 4 effects. Social sector employment is estimated by examining four consecutive quarters in the fourth full year post graduation, and requires a graduate to have been employed in educational services, or government in at least one of these quarters. Sample is limited to graduates who are not still enrolled and who have positive earnings in at least one quarter of the focal year.

Table 11. Institutional Fixed Effects: Cumulative UI Receipt, Year 4 Post-Graduation
Subbaccalaureate institutions

Institution Code	Raw Mean	Model (1)	Model (2)	Model (3)	Model (4)
		<i>No Adjustments</i>	<i>Adj. for student chars.</i>	<i>Adj. for (2) plus zip-level controls</i>	<i>Adj. for (3) plus majors at entry</i>
camp_01	0.21	-0.82	-0.86	-0.78	-1.30
camp_02	0.13	0.99	0.96	0.95	0.80
camp_03	0.15	0.62	1.01	0.88	1.16
camp_04	0.15	0.62	0.40	0.37	0.12
camp_05	0.12	1.28	1.44	1.19	1.34
camp_06	0.17	0.05	0.72	0.60	0.67
camp_07	0.12	1.11	1.55	1.58	1.34
camp_08	0.15	0.45	0.47	0.18	0.20
camp_09	0.20	-0.65	-0.66	-0.73	-0.29
camp_10	0.20	-0.75	-0.62	-0.47	-0.86
camp_11	0.19	-0.33	-0.41	-0.19	-0.37
camp_12	0.11	1.51	1.37	1.30	1.36
camp_13	0.13	0.99	0.94	0.58	0.43
camp_14	0.18	-0.21	-0.34	-0.24	-0.47
camp_15	0.22	-1.05	-1.11	-0.94	-0.68
camp_16	0.18	-0.18	-0.21	-0.22	-0.34
camp_17	0.17	-0.07	-0.14	-0.12	-0.32
camp_18	0.22	-1.13	-1.20	-1.22	-0.91
camp_19	0.21	-0.98	-0.93	-1.01	-0.91
camp_20	0.19	-0.37	-0.25	-0.33	-0.26
camp_21	0.08	2.08	1.82	2.17	2.45
camp_22	0.16	0.22	0.35	0.07	0.15
camp_23	0.18	-0.11	-0.46	-0.22	-0.34
camp_24	0.22	-1.13	-1.32	-1.24	-1.56
camp_25	0.14	0.80	0.38	0.77	0.72
camp_26	0.19	-0.35	-0.30	-0.45	-0.13
camp_27	0.29	-2.68	-2.40	-2.60	-2.13
camp_28	0.17	0.10	-0.19	0.11	0.12
Institution-level mean	0.17	0.00	0.00	0.00	0.00
Institution-level std. dev.	0.04	1.00	1.00	1.00	1.00
<i>Correlations btw. metrics</i>					
Model 1		1.00			
Model 2		0.97	1.00		
Model 3		0.98	0.98	1.00	
Model 4		0.95	0.96	0.97	1.00
Observations	22,467	22,467	22,467	22,467	22,467

Notes: Institutional fixed effects from Models 1-4 are reverse-coded so that lower rates of UI receipt correspond to more positive standardized FE. Cumulative UI receipt is measured as the percent ever receiving UI, or other unemployment compensation, by the end of the fourth full year post graduation. To limit bias from out-of-state mobility, sample is limited to those with positive earnings in at least one quarter of the focal year.

Table 12. Correlations of Adjusted Institution-Level Metrics*SUBBACCALAUREATE INSTITUTIONS*

Correlations	Subba 3yrs	Ftemp 4yrs	Earn 4yrs	SS sec 4yrs	UI 4yrs
Subba 3yrs	1.000				
AA 3yrs	0.646***				
LTC 3yrs	0.467**				
STC 3yrs	0.744***				
Trans 3yrs	0.376**				
Ftemp 4yrs	-0.527***	1.000			
Earn 4yrs	-0.533***	0.818***	1.000		
SS sec 4yrs	0.183	-0.231	-0.026	1.000	
UI 4yrs	-0.330*	0.378**	0.488***	0.222	1.000
Avg diff. vs. BA metric (sd's)	0.00	1.35	1.45	1.06	1.35
Avg diff. vs. earnings metric (sd's)	1.45	0.47	0.00	1.20	0.80

Table 13. Correlations of Adjusted LM Metrics Over Time
SUBBACCALAUREATE INSTITUTIONS

A. Full-time employment proxy

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.977***	1.000		
Year 4	0.920***	0.930***	1.000	
Year 7	0.887***	0.873***	0.890***	1.000

B. Conditional earnings

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.971***	1.000		
Year 4	0.960***	0.975***	1.000	
Year 7	0.883***	0.923***	0.933***	1.000

C. Social sector employment

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.937***	1.000		
Year 4	0.918***	0.955***	1.000	
Year 7	0.841***	0.856***	0.940***	1.000

D. UI Receipt (inverse)

	Year 1	Year 2	Year 4	Year 7
Year 1	1.000			
Year 2	0.894***	1.000		
Year 4	0.798***	0.853***	1.000	
Year 7	0.712***	0.779***	0.942***	1.000

Table 14. Correlations Between Adjusted and Unadjusted Metrics*SUBBACCALAUREATE INSTITUTIONS*

Metric (Standardized)	1 Year	Year 2	Year 4	Year 7
Ftemp	0.941***	0.931***	0.904***	0.913***
Earn	0.912***	0.898***	0.840***	0.815***
SS sec	0.928***	0.955***	0.947***	0.922***
UI	0.950***	0.947***	0.947***	0.939***

Figure 1. Adjusted Institution-Level Metrics (Standardized)

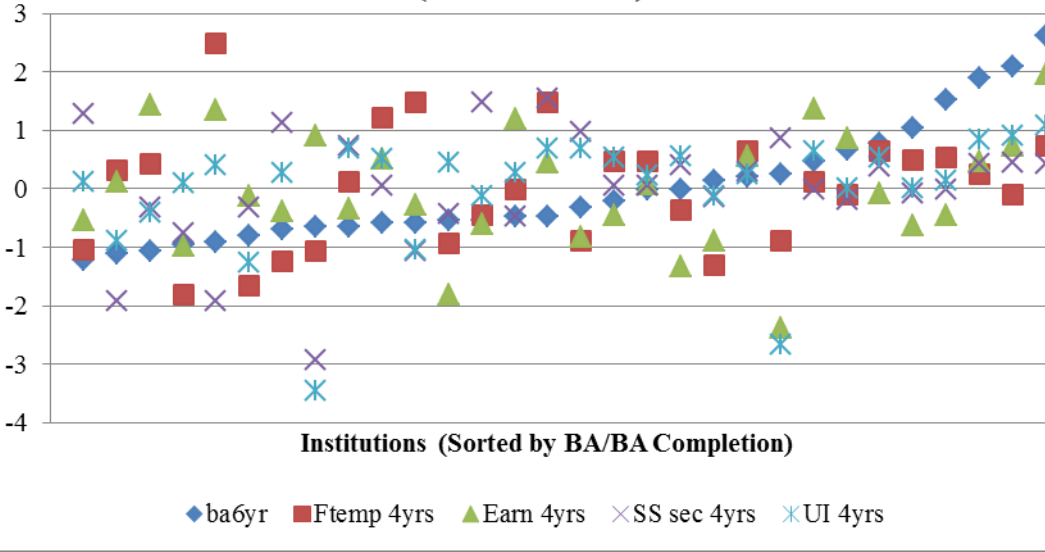


Figure 2. Alternative Earnings Metrics (Standardized)

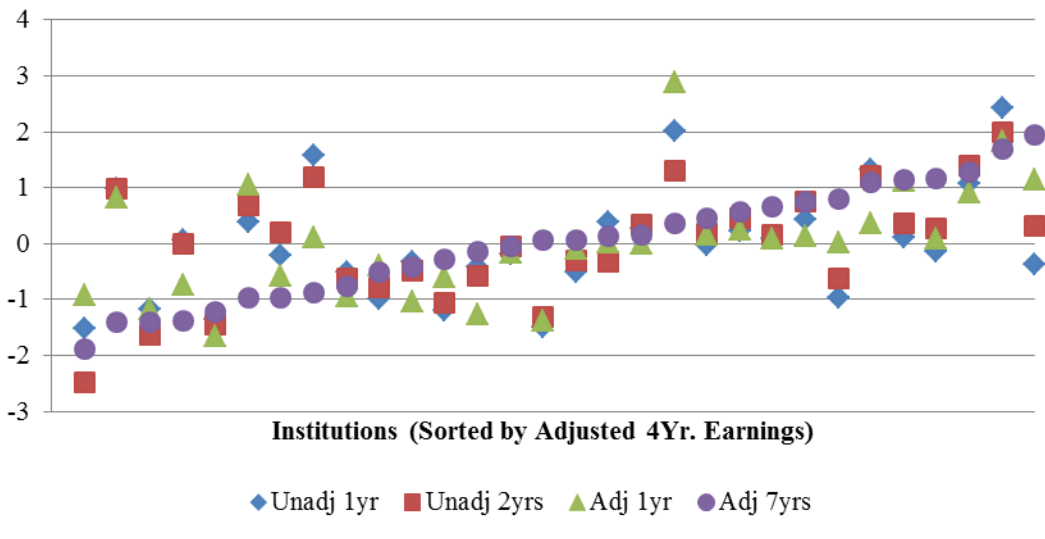
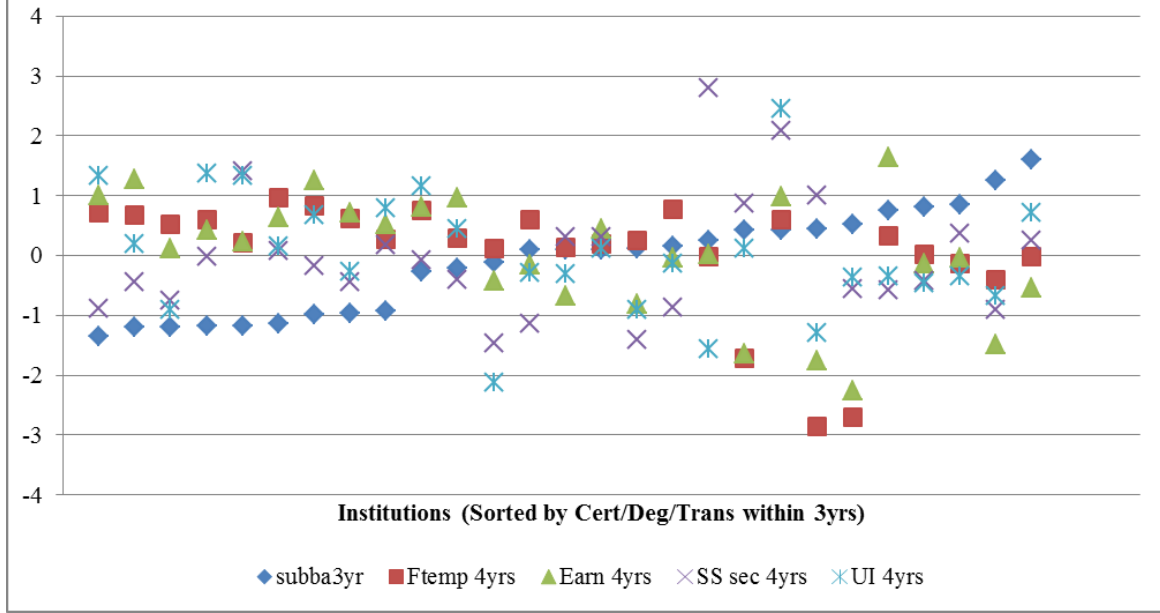


Figure 3. Adjusted Metrics for Sub-BA Institutions



State	Sector Using LMOs	Mandatory v. Optional	Funding Linked to Overall Performance & Role of LMOs	Employment Data source	Sample for LMO Metrics	LMO Metrics
FL	Two- and four-year institutions	Mandatory	\$200 million linked to PBF in 2016, half new funds, half reallocated from base fund. Both universities and colleges have mandatory metrics based on LMOs, though specific metrics vary by sector and LMOs are among several outcomes considered.	Unemployment Insurance (UI) data	Graduates	Percent of bachelor's graduates employed and/or continuing their education further one year after graduation; Median average full-time wages one year after graduation; job placement; average full-time wages compared to entry level wages in local service area.
TX	State technical colleges	Mandatory	100% of state funding recommendations (from THECB to state legislature) are based on LMO.	Unemployment Insurance (UI) data	Graduates AND those who leave TSTCS for at least two years and show up in state workforce.	THECB "value-added" measure is based on five years of post-enrollment earnings, minus the minimum wage.
KS	Two-year institutions	Optional	New state funds; LMOs are options but not mandatory and vary by institution (at least 8 are currently using)	Unemployment Insurance (UI) data	Graduates	Percent of students employed; % employed in field; Wages of students hired; starting wages
AK	Two-year institutions	Optional	Five percent of funding in the 2012-2013 school year, and increasing by 5% increments until capped at 25% during the 2017-2018 school year. The remaining 75 percent of funding will be based on enrollment and institutional needs. LMO metric is optional, not mandatory, unclear if any campus is using.	Unemployment Insurance (UI) data	Graduates	Number of completers that obtain employment (timeframe unclear)
LA	Two- and four-year institutions	Optional	15% of base appropriations—institutions can also receive permission to raise tuition by 10% without legislative approval; LMOs are options but not mandatory and not clear that any institutions are currently using.	?	Graduates	Employment of degree and certificate earners
MN	Two-year institutions	Mandatory	5% of base funding is reserved until institutions meet three out of five performance goals.	Graduate follow-up survey data	Graduates	Percent employed in program-related job during the year after graduation.
WI	Technical colleges	Mandatory	10% of base funding will be at stake during FY2014-15. The amount of performance funding will increase by 10% increments until reaching 30% in FY2016-17	Graduate follow-up survey data	Graduates	Placement rate of students in jobs related to students' programs of study
TN	Two-year institutions	Mandatory	After a base amount is set aside for operational support, 100% of state funding is allocated based on institutional outcomes	?	Graduates	Job Placement within a related field through June 30 of the year following graduation.

Source: NCSL (2015), THECB (2013), FLBOG (2016), and state agency websites.