Economics of Education Review 42 (2014) 64–77

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev

Measuring value-added in higher education: Possibilities and limitations in the use of administrative data☆

Jesse M. Cunha [a,*], Trey Miller [b]

[a] Naval Postgraduate School, Graduate School of Business and Public Policy, 555 Dyer Road, Ingersoll Hall, Monterey, CA 93943, United States
[b] RAND Corporation, 1776 Main Street, Santa Monica, CA 90401, United States

ARTICLE INFO

Article history:
Received 12 November 2012
Received in revised form 10 June 2014
Accepted 11 June 2014
Available online 18 July 2014

JEL classification:
I22
I23
I28

Keywords:
Higher education
Value-added
Administrative data
Graduation
Persistence
Earnings

ABSTRACT

This paper develops a general methodology for measuring the value added of institutions of higher education using commonly available administrative data. Our approach recognizes the data limitations and selection problems inherent in higher education, and highlights the challenges these issues pose for education policy. Combining information from different administrative sources in the state of Texas, we follow the universe of Texas college applicants from the time of application (pre-enrollment) through public college and into the labor market. In specifications that do not control for selection, we find large, significant differences across colleges in terms of persistence, graduation, and earnings; however, these differences decrease substantially when we control for selection. In light of the growing interest in using value-added measures in higher education for both funding and incentivizing purposes, our methodology offers unique evidence and lessons for policy makers.

"Student achievement, which is inextricably connected to institutional success, must be measured by institutions on a 'value-added' basis that takes into account students' academic baseline when assessing their results. This information should be made available to students, and reported publicly in aggregate form to provide consumers and policymakers an accessible, understandable way to measure the relative effectiveness of different colleges and universities."
Quote from "A Test of Leadership," the 2006 Report of the U.S. Department of Education (the Spellings Commission) on Higher Education

☆ We thank the Texas Higher Education Coordinating Board for its support of this project, particularly Susan Brown, David Gardner, and Lee Holcombe. We also thank Doug Bernheim, Eric Bettinger, Charles Clotfelter, David Figlio, Rick Hanushek, Doug Harris, Caroline Hoxby, Giacomo De Giorgi, Paco Martorell, Sarah Turner, and participants in the Bill & Melinda Gates Foundation/HCM Strategists "Context for Success" Project for helpful comments. This work updates and extends two earlier reports: a 2008 white paper written for the Texas Higher Education Coordinating Board entitled *Quantitative Measures of Achievement Gains and Value-Added in Higher Education: Possibilities and Limitations in the State of Texas* and a 2011 white paper written for the Bill & Melinda Gates Foundation/HCM Strategists entitled *Measuring Value-added in Higher Education*.

* Corresponding author. Tel.: +1 650 492 0381.
E-mail addresses: jcunha@nps.edu, jessecunha@gmail.com (J.M. Cunha), tmiller@rand.org (T. Miller).

http://dx.doi.org/10.1016/j.econedurev.2014.06.001

1. Introduction

Recent years have seen mounting pressure on colleges and universities to measure and disseminate the value that they are adding to their students (see, e.g., Harnisch, 2011). This outcomes-based culture recognizes foremost the need for measures of value added that capture the causal influence of institutions on their students, which must take into account the fact that students enter college with different academic backgrounds (Spellings, 2006). Many government agencies – of both states in the U.S. and other countries – are using or are considering using quantitative measures of institutional performance to incentivize achievement and target funding (Dougherty, Natow, & Vega, 2012; Jongbloed & Vossensteyn, 2001). For example, in the U.S., the state of Tennessee is currently using a performance-based funding formula in higher education, and the state of Texas is considering using one (Wright, Fox, Murray, Carruthers, & Thrall, 2012); in the United Kingdom, there is continual interest in using quantitative performance indicators for public colleges and universities for funding purposes (Johnes, 2012). However, there is a lack of research that can guide policymakers towards an optimal policy.

In this paper, we discuss the unique challenges of measuring value added in higher education and explore the possibilities and limitations of using commonly available student-level administrative data as the basis for such measures. It proves useful to contrast the higher education environment with that of primary and secondary education, a sector which is broadly characterized by the use of yearly standardized test scores as the basis of value-added measures. While there is evidence that such test-based value-added measures can indeed capture differential performance of institutions and teachers (Kain & Staiger, 2008; Klein, Kuh, Chun, Hamilton, & Shavelson, 2005), several differences render the wholesale importation of the K–12 model to higher education impractical.

First, year-on-year standardized tests are not generally administered in higher education.[1] In their stead, we must consider other quantitative outcomes of the higher education process that are available in administrative databases. Three such outcomes of interest are persistence rates, graduation rates, and post-college earnings.

Second, these available non-test outcomes are either only observed once (persistence or graduation) or only observed post-enrollment (earnings). As such, we cannot use a within-individual differencing estimator – an estimator which can be extremely useful in order to isolate the influence of specific factors in the education process, such as teachers and schools, separately from pre-existing student ability.[2]

Third, students deliberately and systematically select into colleges.[3] Combined with the lack of pre-enrollment outcome measures, this selection problem makes it difficult to attribute student outcomes to the effect of the college attended separately from the effect of pre-existing characteristics such as motivation and natural ability.

Fourth, college students intentionally specialize their instruction, and institutions emphasize discipline-specific knowledge (i.e., major specific knowledge). Such specialization calls for outcome measures that are comparable across students with a wide range of learned abilities. In this respect, standardized tests of general skills may not be the optimal outcome measure.

Reflecting the unique context of higher education and the availability of data, we propose a simple methodology that provides estimates of the relative value added of individual institutions: a student-level regression that explains the variation in the outcome of interest through (i) observable differences in pre-enrollment student characteristics, (ii) unobserved differences in students' preferences for schools and schools' preferences for students, captured by a student's application and acceptance profile, and (iii) fixed effects for the college at which a student is enrolled. This model yields average differences in conditional outcomes across colleges, or relative value-added measures; and these measures can be considered causal value-added estimates to the extent that pre-enrollment student characteristics and application/acceptance profiles control for differential selection into colleges.

We implement this methodology using rich administrative records from the state of Texas, developing value-added estimates for the state's 30 traditional four-year public colleges. Texas has one of the most-developed K–20 data systems in the nation and thus provides an ideal setting to demonstrate the incremental benefits of using various student-level data sources to correct for selection, while at the same time demonstrating the potential bias that can result by not correcting for selection.

Our analysis shows that there are large mean differences in outcomes across public colleges prior to controlling for pre-existing student characteristics. For example, the unconditional mean difference in earnings between Texas A&M University and Texas Southern University – the institutions with the highest and lowest unconditional earnings, respectively – is 78 log points. Perhaps not surprisingly, our analysis confirms that value-added measures change considerably upon controlling for pre-enrollment student characteristics. Continuing the example, controlling for the largest set of student characteristics

[1] Some standardized tests in higher-education do exist, for example the Collegiate Learning Assessment (CLA) test of general knowledge or the GRE Subject Tests of major specific knowledge; however, they are not to our knowledge administered regularly to all students in an administrative unit (e.g., a state or country).

[2] Certain populations do have labor market experience prior to college enrollment, facilitating student fixed-effect models in labor market earnings. For example, Cellini and Chaudhary (2012) study the returns to private, two-year colleges, and Arcidiacono, Cooley, and Hussey (2008) study the returns to MBA programs.

[3] Sorting undoubtedly occurs in the primary and secondary setting as well (Tiebout, 1956); presumably, however, to a much smaller degree.

and their unique application and acceptance profile, the difference in mean earnings between Texas A&M and Texas Southern decreases to 30 log points. A similar pattern is seen when using persistence and graduation as outcomes, and when comparing amongst various student subgroups.

Our work relates to several different literatures. A number of studies have used standardized test scores to estimate value added in higher education for select groups of students and schools. For example, Klein et al. (2005) administered a test of cognitive skills to a small sample of students from diverse colleges in an effort to validate the usefulness of such tests; Liu (2011) compares the results of two different methodologies that measure value added of institutions using standardized test scores; and Saavedra and Saavedra (2011) use standardized tests administered to a sample of Colombian college students to estimate the value added of a college education. Several papers have also used earnings to estimate value added in higher education. For example, Mallier and Rodgers (1995) propose and estimate a measure that compares earnings between graduates and non-graduates in the United Kingdom, and Rodgers (2007) explores how this methodology can be used to develop performance indicators for schools.

When the outcome in question is earnings, we study the same question as in the literature on the labor market return to college quality (see e.g., Andrews, Li, & Lovenheim, 2012; Behrman, Rosenzweig, & Taubman, 1996; Black & Smith, 2004; Brewer, Eide, & Ehrenberg, 1999; Zhang, 2009). While this literature is mostly concerned with addressing selection into college, our work is unique in estimating returns for individual schools, rather than using an index of college quality or selectivity (such as the school average SAT score).

Furthermore, our work also provides a practical application of the Dale and Krueger (2002) methodology – that controlling for the set of applications and acceptances can help account for unobserved selection into schools – and we show how value-added estimates differ with and without using applications and acceptances as controls. Another recent example of the use of this methodology is Broecke (2012), who estimates returns to college selectivity for students in the United Kingdom.

This work is also related to a strand of literature that uses natural experiments which provide a locally exogenous selection rule into college to overcome the selection problem. For example, Hoekstra (2009) exploits a strict SAT-score cutoff used for admission by a large flagship university to compare the earnings of applicants scoring just above and below the admission cutoff, and Saavedra (2009) exploits a sharp discontinuity in acceptance based on a standardized college entry test in Colombia to measure the effect of attending a better-quality school on college exit scores, employment, and earnings. While attractive, discontinuities of this type are rare, reducing the usefulness of this methodology for policymaking purposes.

This paper proceeds as follows. First, we discuss in Section 2 the value of higher education and potential quantitative indicators of value in this setting. In Section 3, we discuss the empirical challenges in the measurement of value added in higher education, and outline our empirical model. Section 4 describes the context of our empirical sample and the data we use. Section 5 presents results of the analysis, and Section 6 concludes.

2. Defining value in higher education

2.1. Defining the value of higher education to students

Colleges aim to produce a wide range of benefits for students.[4] Perhaps foremost, colleges aim to increase students' individual utility: A college education imparts knowledge and skills that increase students' economic productivity and hence their wages, and higher wages allow for an expanded budget set of consumption-leisure bundles.

Colleges also foster non-pecuniary benefits, such as offering greater choices in the type of work one performs, an improved ability to make informed life decisions about marriage, health, and parenting, and even perhaps an increased sense of general happiness (Oreopoulos & Salvanes, 2011). Furthermore, knowledge in and of itself may have utility value for some – the "consumption value" of knowledge.

Finally, colleges foster positive externalities accruing to society at large, benefits that students will enjoy throughout their lives. These externalities operate through increasing returns to knowledge, returns that are increasing in the proximity of other knowledgeable people, such as in firms (Acemoglu & Angrist, 2001) or cities (Moretti, 2004). Similarly, well-informed citizens are more likely to contribute to public goods that improve the functioning of civic society, through acts such as voting and supporting free speech (Dee, 2004).

These wide-ranging goals of higher education suggest a role for multidimensional performance measures. Our goal in this paper is not to suggest the particular outcome measures that should be used to measure value added in college; rather, we argue that a full appraisal of institutional performance at the postsecondary level likely requires a set of indicators that proxy for the various dimensions of institutional performance, and this set of indicators is necessarily limited by data availability.

2.2. Student-level outcomes to measure institutional performance

In this light, we summarize potential student-level outcomes that could be used to measure dimensions of institutional performance, noting both their advantages and disadvantages, as well as their availability in common administrative databases.

2.2.1. Standardized tests

Test scores are a succinct and practical way to assess knowledge; compared year-to-year, they provide a useful measure of knowledge gains. Indeed, standardized tests

[4] For a more in depth survey of research on the affect of college on students, see Pascarella and Terenzini (2005).

have been developed to measure both general and specific knowledge in higher education. Tests of general knowledge in higher education include, amongst others, the Collegiate Learning Assessment (CLA), ACT's Collegiate Assessment of Academic Proficiency (CAPP), and ETS's Measure of Academic Proficiency and Progress (MAPP). Examples of tests of specific knowledge include ETS's Major Specific Tests and GRE subject tests for certain disciplines. Aside from certain licensure exams, such as for accountants or nurses, neither general nor specific tests are currently applied to a student body in the U.S. as general practice.

While the use of these tests in the future as a basis for measuring value added is appealing, one issue that will likely complicate their use is that many students drop out of college, and differential attrition across schools could make it difficult to use school average test scores of general skills in a meaningful way. A separate issue is that it is unclear how to compare gains in knowledge across disciplines.

2.2.2. Grade point average (GPA)

GPA or individual course grades are a natural consideration as outcomes of student performance, and they are generally included in administrative databases. However, grades are commonly normalized within a class to a predetermined distribution; as such, it is difficult to compare them within schools (specifically, across majors) or across schools. For example, consider comparing an "A" in calculus courses at two different colleges, when it is known that one teaches a more rigorous course.

2.2.3. Graduation and persistence

Whether a student persists through college and ultimately graduates is certainly of interest to policymakers. One appeal of graduation rates as an indicator of the value obtained in college is that the degree granting process standardizes to some extent the knowledge and skill levels deemed necessary to receive that degree. What is unclear, however, is the extent to which obtaining a college degree informs us of the *quality* of the knowledge and skills the student acquired while in school. Nonetheless, graduation and persistence data are commonly available in administrative databases, and, in fact, completion rates are currently incorporated in several U.S. states' performance-based funding models (Midwestern Higher Education Compact, 2009).

One limitation of graduation rates is that they are observed many years after a student initially enrolls in school, which introduces a long time lag between any changes in policy and the observation of that policy's effect. Naturally, one-year persistence rates overcome this limitation and allow for more immediate policy feedback.

2.2.4. Wages/earnings

Wages are an attractive measure of achievement in college for several reasons. First, a simple theory of competitive labor markets predicts that the wage rate equals a worker's marginal product: The more a worker adds to the value of a final product, the higher her wage will be. Higher wages signal high productivity, and productivity is increased through knowledge and skills, which can be learned in college.

Second, unlike tests, the wage rate aggregates the influence of both general and specific skills in accordance with their importance for economic productivity. For example, a nurse's wage reflects general and specific skills, say communication skills and wound-care skills, while a civil engineer's wage reflects her general and specific skills, say problem-solving and design skills. Therefore, in a perfectly competitive labor market, any two workers with the same wage are equally productive despite having different occupations. Finally, wages can be observed both for those who graduate from college and those who do not. Thus, one's wage not only reflects their productivity, but also reflects any possible graduation effect (a "sheepskin" effect).

Labor markets, however, may not be perfectly competitive, and the wage rate becomes less attractive as a measure of value to the extent that market frictions exist. For example, unionized occupations may receive wages in excess of their productivity. Furthermore, several other disadvantages of wages as a measure of the value of education can be identified: if workers are intrinsically motivated, as is often the case in the nonprofit sector, then wages do not fully reflect productivity; wages are only observed for those who work, which excludes, for example, home care; and the social benefits of education, such as civic-mindedness, are likely not reflected in wages.

In addition, we typically do not observe wage rates in administrative databases, but rather, earnings over some time period (a quarter or year). Earnings are less attractive than wages, as they combine both the wage rate and the labor supply decision. Several other disadvantages of the use of earnings are of note. First, earnings data that can be matched at the individual level to educational databases usually come from state unemployment records that only include earnings from jobs that are eligible for unemployment insurance (UI) benefits. These include the majority of earnings from wages and salaries, but exclude other sources, such as self-employment income. UI records also exclude earnings from federal government jobs, including those in the military. UI earnings thus systematically underreport total earnings, and this is a concern if students who attend some institutions are more likely to generate earnings from uncovered sources than are students who attend other institutions.

Second, any state-level earnings database only contains earnings information for jobs held within the state, implying that we can only use such databases to track the earnings of college students who maintain residence in the state. This is a concern if there is differential out-of-state migration of students across colleges and students who are employed out of state have systematically higher or lower earnings than those employed in state. Finally, as with graduation rates, there is a long time lag between enrollment in college and the observation of labor market earnings (usually after graduation).

3. Measuring value added in higher education

Consider a generic educational production function that could be used to estimate differences in student outcomes across schools when the outcome of interest is observed pre-enrollment:

$$Y_{is} = \delta Y_{i,PRE} + \Phi \mathrm{X}_{\mathrm{i,PRE}} + \sum_{s=1}^{n} \beta_s E_s + \epsilon_{is} \quad (1)$$

E_s are a set of indicators for enrollment at various colleges s, Y_{is} is an outcome for student i who attended school s, $Y_{i,PRE}$ is the pre-enrollment value of the outcome measure, and $\mathbf{X_{i,PRE}}$ is a vector of observable pre-enrollment student characteristics. It would be convenient if the coefficients β_s were interpretable as the school-average, over-time change in the outcome that is attributable to the effect of attending the school in question. However, such a causal interpretation can be confounded for several reasons.

First, students may select into colleges in ways that are both unobservable to the analyst and correlated with the outcome of interest, even conditional on $Y_{i,PRE}$ and $\mathbf{X_{i,PRE}}$. For example, students are likely to have private knowledge about their future career goals that is not reflected in their observable actions when applying to schools, and they are likely to select colleges that best advance those career goals. Indeed, this type of self selection into colleges is actively encouraged by guidance counselors, teachers, and parents alike.

Second, pre-college outcome measures, $Y_{i,PRE}$, are often not observed (as is often the case with standardized test scores or earnings) or do not exist (as with indicators of persistence and graduation). Thus, we are forced to estimate:

$$Y_{is} = \Phi \mathrm{X}_{\mathrm{i,PRE}} + \sum_{s=1}^{n} \beta_s E_s + \varepsilon_{is} \quad (2)$$

and we must rely solely on $\mathbf{X_{i,PRE}}$ to capture unobserved selection effects that may be correlated with outcomes and school choice. Furthermore, the lack of pre-enrollment outcome measures implies that β_s in Eq. (2) reflects the relative differences in outcomes across schools (as opposed to within school changes over time).[5]

Third, the outcome of interest may not be observed for all students. For example, students may be absent on the day of a standardized test; earnings may not be observed if a student moves out of the state in which they attended school; or students may transfer to different colleges not covered by the database at hand. Differential observability of outcomes may bias estimates of β_s.

Thus, barring pre-enrollment data on the outcome of interest, a causal interpretation of β_s relies on both $\mathbf{X_{i,PRE}}$ adequately controlling for selection into colleges and ensuring that outcomes are observed for all enrollees. In practice, we argue that $\mathbf{X_{i,PRE}}$ should include all observable characteristics available to the researcher that influence the choice of college and the outcome of interest. Typical characteristics available in administrative datasets include gender, ethnicity, high school course achievement (courses and grades), and standardized test scores (e.g., SAT scores).

Certainly, some observable variables available to the researcher have been affected by the student's choice of college; for example, the choice of major or participation in sports or activities. If such variables are included in $\mathbf{X_{i,PRE}}$, estimates of β_s would exclude the direct effect of these variables on the outcome of interest. It may certainly be of interest to estimate models conditioning on post-enrollment choices (such as one's major), but such a model will not yield the comprehensive value added of colleges. Thus, variables that could be affected by college choice should rightfully be considered as outcomes rather than covariates.

3.1. Controlling for selection on observables

One set of conditioning variables deserves particular discussion: the set of schools to which a student applied and was accepted. Dale and Krueger (2002) argue that conditioning on this profile of applications and acceptances captures information about the selection process that would otherwise be unobserved by the researcher. Specifically, they argue that students use private information to optimize their college application decisions, and colleges use information unobservable to researchers (such as the quality of written statements and letters of recommendation) to make acceptance decisions. Thus, students with identical college application and acceptance profiles are likely to be similar on both observable and unobservable dimensions.

Of course, even controlling for application and acceptance profiles is unlikely to perfectly describe selection into college. In particular, this identification strategy essentially compares students with the same observable characteristics who applied to and were accepted at the same set of colleges, yet decided to attend different colleges. It would be naïve to believe that this ultimate choice of which school to attend was uncorrelated with the outcomes of interest.

4. Data and sample

4.1. Higher education in Texas

We demonstrate the applicability of this value-added model using rich administrative data from the state of Texas that tracks students from high school, through college, and into the labor force. For at least two reasons, Texas is an excellent setting to estimate these models. First, it is a large and diverse state that closely mirrors the demographic and socio-economic make-up of the U.S. population. Second, the vast majority of Texans attend college in state; for example, 88 percent of Texas high school graduates who enrolled at a four-year college in 2002 enrolled at a Texas public college (Snyder, Dillow, & Hoffman, 2007). This second characteristic helps mitigate sample selection issues that would likely be far more problematic in a smaller state where students tend to migrate out of state either for college or to work after graduating.

[5] An alternative specification that includes high school graduates who never attended college would recover the absolute effects of attending a particular college (over not attending college); however, this specification is only available for outcomes that are observed even if one does not attend college, such as earnings.

Our analysis focuses on the 30 traditional public four-year universities in the state during our sample period, which includes the entering first-year cohorts of the 1998–2002 academic years.[6,7] These traditional four-year universities can be grouped into three main categories: flagship institutions, emerging research institutions with statewide draw, and regionally focused institutions.

The two flagship institutions, the University of Texas at Austin and Texas A&M University at College Station, are nationally recognized Ph.D.-granting research institutions that offer a comprehensive set of majors and draw a diverse range of students from across the state and nation. The six emerging research institutions are major Ph.D.-granting institutions that draw their students mainly from within the state and also offer a comprehensive set of majors. These include Texas Tech University, University of Texas at Arlington, University of Texas at San Antonio, University of Texas at El Paso, University of Texas at Dallas, University of Houston, and University of North Texas. The remaining 22 public four-year institutions generally do grant Ph.D.s, but they vary in major and course offerings and tend to attract students from within their geographical region. Online Appendix Table 1 lists all of the schools included in our analysis and summarizes several additional institutional details.

The application and admissions process at Texas four-year public colleges is similar to the process in most other states. Students may submit application packages to as many institutions as they desire, and a common application is accepted at all Texas public institutions. The common application includes the high school transcript and some basic demographic information. Students separately submit their entrance exam scores (e.g., the SAT) via the relevant testing agency. Certain institutions also require institution specific material, such as one or more essays and/or letters of recommendation.

Institutions consider a number of factors when deciding whether to admit a student, but most weight is placed on the applicant's high school GPA and entrance exam score. Typically only the selective institutions place much weight on other factors including student extracurricular activities during high school, essays or letters of recommendation.

4.2. Data

Our data come from various administrative agencies in Texas, and are housed securely by the Texas Higher Education Coordinating Board (THECB). THECB collects information on all individuals who apply to or attend any college in the state, including indicators of which school(s) a student applied to and whether or not she was accepted, the number of credit hours a student was enrolled for each semester, and whether or not she graduated.

To this collegiate database, we merge three other data sources at the individual level using students' Social Security Numbers (SSN). Data on Texas public high school graduates is provided by the Texas Education Agency (TEA), which includes courses taken, programs participated in, an indicator of belonging to a low-income family (eligibility for free school lunch), a TEA-defined indicator for being at risk of not graduating high school,[8] and the student's gender and race/ethnicity. SAT score data is purchased from the College Board, as is survey data collected at the time of the test, which includes self-reported GPA, class rank, planned educational attainment, and father's and mother's education level. Finally, quarterly earnings data come from the state's Unemployment Insurance (UI) benefits program, which is available for all workers employed in benefits-eligible jobs in the state.[9] While some employment in Texas is excluded from the Texas UI database (e.g., self-employment, certain exempt occupations), there is evidence that the vast majority of employment is included (see Stevens (2002) for details).

We note that this database contains all of the data that is observed by university selection committees, except for entrance essays, letters of recommendation, and a student's set of extracurricular activities, when these factors are required.

4.2.1. Outcome variables

Using this data, we define several outcome variables. A student is defined as graduating from college if she receives a bachelor's degree at any public or private four-year college in Texas by the end of the sixth academic year after graduating from high school. Similarly, we define a student as persisting into year two of college if she completes at least 30 credit hours at any public or private four-year college in Texas by the end of the sixth academic year after graduating from high school.

Note that these definitions allow for transfers within the Texas public school system, yet attribute the effect of graduating/persisting to the school of initial enrollment (as long as that initial enrollment was at a public four-year college). This choice of definitions is consistent with the principle espoused previously that any outcome subsequent to initial enrollment is rightfully attributable to the effect of the first school attended.

[6] We do not include five non-traditional four-year public universities (Sul Ross University Rio Grande College, University of Houston-Clear Lake, University of Houston-Victoria, University of North Texas Dallas, and Texas A&M Texarkana) and one university that opened in the middle of our sample period (UT Brownsville).

[7] Due to a lack of data, we do not consider private colleges nor two-year colleges, although in principle the methodology we propose can be equally applied to any class of institutions.

[8] The indicator for being at risk of not graduating high school takes a value of one if a student has any of a set of characteristics identified by the TEA as predicting high school drop-out; examples include failing at least two courses in the basic high school curriculum, being retained within grade, being pregnant or a parent, having been expelled, or being on parole. For a comprehensive list of characteristics that define the "at risk" indicator, see the TEA data standards at http://ritter.tea.state.tx.us/trex/datastds/.

[9] The unemployment insurance records in Texas – as well as many other states – also indicate the industry of the employer in each job held by an individual worker. Unfortunately, this information is not particularly useful for our purposes, as it is impossible to distinguish between occupations within an industry (for example, accountants, engineers, and custodians are not distinguishable within the oil industry).

Furthermore, the choice of time horizon for persistence and degree completion involves a tradeoff: a short horizon will exclude students with longer-than-normal college careers, while a long horizon introduces a large lag between the action of institutions and the measurement of student response (i.e., graduating), thus reducing the utility of value-added measures for policymaking purposes. While a six-year completion window is consistent with much of the academic literature, policymakers must necessarily weigh these competing demands in light of the context at hand.

Annual earnings, converted to 2010 dollars using the CPI-U, are calculated as the sum of quarterly earnings reported to the state in the eighth calendar year after graduating from high school. As with graduation and persistence, this choice of time horizon involves a tradeoff: on the one hand, a longer time horizon ensures that all students have had sufficient time to complete college and be absorbed by the labor market; on the other, it also increases the lag time between the actions of the institution and the measurement of the response.

An important limitation of using UI earnings data is that we do not observe earnings for students who migrate out of state, making it impossible to distinguish between those individuals who do not work, are not working in a benefits-eligible job, or have moved out of state. Importantly, this limitation is not specific to Texas, as no state unemployment insurance commission tracks earnings across state lines. This issue of missing data can be problematic for our proposed value-added methodology if there is a correlation between choice of school and having zero UI earnings. Given this limitation, two options are available: (i) impute missing UI data with zero earnings, or (ii) treat missing UI data as missing when constructing our measure of earnings. We choose the latter approach, acknowledging that the choice of approach is necessarily context specific and will dictate the interpretability of resulting value-added estimates.

4.3. Sample and summary statistics

Our empirical analysis uses data for students who graduated from any Texas public high school between 1998 and 2002 and subsequently enrolled in a Texas public four-year college. Note that this sample excludes the approximately 21 percent of enrollees from several groups for whom we do not have pre-enrollment data: students that were home schooled, that attended a private high school in Texas, or that attended any high school out of state.

Table 1 summarizes the data, for both the sample of all enrollees and only enrollees with non-zero UI earnings. Looking first at the sample of all enrollees, annual earnings (amongst those with reported earnings) average about $37,000 dollars, 60 percent graduate within six years of enrolling, and 84 percent persist into year two of college within six years of enrolling. We do not exclude students that are enrolled in graduate school, or are still enrolled in an undergraduate degree. While it may be desirable to exclude these students in certain contexts, we include them in keeping with the principle that the choice of enrolling in graduate school or not graduating may be influenced by one's undergraduate experience.

The majority of the sample (60 percent) is white, although there is a significant Hispanic population. Eighteen percent of the sample is low income, as indicated by their eligibility for free lunch, about a third were at risk of not graduating high school, and the average composite SAT score of 1042 points is, not surprisingly, close to the nationwide average (The College Board). The sample excluding enrollees with zero earnings is observably similar to the full sample, reflecting the heterogenous composition of the zero-earnings sample, which likely includes both low- and high-skilled individuals.

Note that SAT data is missing for about 23 percent of the full sample, despite the fact that all of the colleges in our sample required SAT scores for admission. It is likely that these students incorrectly reported their SSN to the College Board, and we include an indicator for missing SAT data in the regression analysis below.

4.3.1. Application and acceptance profiles

Table 2 contains the distribution of the number of applications students made to the 30 schools in our sample. The vast majority of students (63 percent) applied to only one school (and application and acceptance data provide no extra useful information for these students). Also, very few students applied to more than three schools. In order to characterize the sets of schools that students applied to and were accepted at, we create indicator variables as follows. For students who applied to two or three schools, we create indicators for the unique combinations of applications and acceptances, yielding 2350 groups.[10] For computational feasibility, we create one indicator variable for students who applied to four or more schools. Finally, for about five percent of students, we do not observe an application to any school. This missing information is likely due to mis-reporting of a student's SSN (our matching variable) by colleges to the THECB, and we include an indicator in regressions below for students with no application data.

5. Results

We present estimates of Eq. (2) for three outcomes: persistence, graduation, and earnings. For consistency, we exclude the indicator for enrollment at Texas A&M University in all models (Texas A&M graduates have the highest unconditional earnings); thus, estimates of β_s are interpretable as mean differences in the outcome relative to Texas A&M, conditional on the included covariates.

[10] For students who applied to two schools, there are $C_2^{30} \times 3 = 1305$ possible combinations of applications and acceptances: for each of the C_2^{30} unique combinations of schools applied to, there are three possible acceptance outcomes (accepted only at one school or the other, or accepted at both schools). Using a similar logic, there are $C_3^{30} \times 6 = 24{,}360$ possible combinations of applications and acceptances for students who applied to three schools. Amongst these 25,665 possible combinations, only 2350 contain two or more students, a result stemming largely from the fact that students in Texas tend to not apply to regional schools outside of their home area.

Table 1
Summary statistics of outcome variables and selected control variables.

	All enrollees			Enrollees with non-zero earnings		
	Mean	(s.d.)	Obs.	Mean	(s.d.)	Obs.
Outcomes						
Annual earnings ($)	36,474	(41,910)	169,239	36,474	(41,910)	169,239
Graduated	0.60	(0.49)	217,733	0.61	(0.49)	169,239
Persisted into year 2 of college	0.84	(0.37)	217,733	0.85	(0.36)	169,239
Student demographics						
Black	0.12	(0.33)	217,733	0.13	(0.34)	169,239
Hispanic	0.21	(0.41)	217,733	0.22	(0.42)	169,239
White	0.60	(0.49)	217,733	0.59	(0.49)	169,239
Other race	0.06	(0.25)	217,733	0.06	(0.23)	169,239
Male	0.45	(0.50)	217,733	0.44	(0.50)	169,239
High school variables						
Courses and programs						
English as a 2nd language	0.13	(1.07)	217,733	0.13	(1.05)	169,239
Gifted and talented program	0.38	(1.13)	217,733	0.37	(1.11)	169,239
Calculus	0.23	(0.44)	217,733	0.22	(0.43)	169,239
Pre-Calculus	0.24	(0.43)	217,733	0.25	(0.43)	169,239
Algebra 2	0.07	(0.26)	217,733	0.07	(0.26)	169,239
Biology	0.02	(0.13)	217,733	0.02	(0.13)	169,239
Chemistry	0.02	(0.14)	217,733	0.02	(0.14)	169,239
Physics	0.05	(0.22)	217,733	0.05	(0.22)	169,239
Student status						
Eligible for free lunch	0.18	(0.38)	217,733	0.18	(0.39)	169,239
At risk of not graduating	0.29	(1.11)	217,733	0.29	(1.09)	169,239
SAT						
SAT composite score	1042.1	(181.2)	167,581	1031.2	(178.2)	129,765

Notes: (1) "All enrollees" sample includes all graduates of Texas public high schools between 1998 and 2002 who ever enrolled in a Texas public college. "Enrollees with non-zero earnings" restricts the "All enrollees" sample to those who had strictly positive earnings reported in state Unemployment Insurance records eight years after graduating from high school. (2) Annual earnings are calculated as the sum of quarterly earnings as recorded in Texas state Unemployment Insurance records in the eighth calendar year after graduating from high school, converted to 2010 dollars using the CPI-U. If no earnings are reported, the observation is treated as missing (see text). (3) Graduating is defined as the student receiving a bachelor's degree at any public or private four-year college in Texas by the end of the sixth academic year after graduating from high school. (4) Persistence into Year 2 of college is defined as the student completing at least 30 credit hours at any public or private four-year college in Texas by the end of the sixth academic year after graduating from high school.

5.1. Outcome: earnings

We begin by examining how conditional earnings change upon controlling for successively more comprehensive sets of pre-college enrollment covariates. Table 3 displays β_s from various specifications of Eq. (2) in which the outcome is the logarithm of earnings eight years after graduating from high school. Column 1 conditions only on high-school graduation year fixed effects, which absorb any aggregate differences in earnings across years, such as the effect of inflation or general macroeconomic shocks. Thus, the coefficients in column 1 can be interpreted (up to an approximation) as the average percentage differences in unconditional earnings of enrollees at various Texas colleges, relative to enrollees at Texas A&M. For example, UT Dallas enrollees earned on average 12 percent less than Texas A&M enrollees, while Texas Southern enrollees earned on average about 118 percent less than Texas A&M enrollees.[11] Perhaps not surprisingly, the relative differences in unadjusted log earnings correlate highly with common perceptions of college "quality" published in the popular press such as *U.S. News & World Report* or *Barron's Magazine*.

Furthermore, note that all of the estimates in column 1 are significantly different from zero at the one percent level; as such, we are fairly confident that average

Table 2
Distribution of the number of applications to Texas four-year public colleges.

Number of applications	Frequency	Percent
0	11,812	5.42
1	136,680	62.77
2	43,377	19.92
3	17,081	7.84
4	5814	2.67
5	1879	0.86
6	588	0.27
7	221	0.1
8	113	0.05
9	54	0.02
10 or more	114	0.04
Total	217,733	100

Note: Sample includes all graduates of Texas public high schools between 1998 and 2002 who ever enrolled in a Texas public college.

[11] $exp(0.78) = 2.18$.

Table 3
Earnings differences of enrollees at Texas four-year public colleges. Conditioning on various subsets of observable covariates.

Outcome =	ln(earnings) (1)		ln(earnings) (2)		ln(earnings) (3)		ln(earnings) (4)		ln(earnings) (5)	
UT Tyler	−0.24**	(0.04)	−0.23**	(0.04)	−0.15**	(0.04)	−0.13**	(0.04)	0.01	(0.08)
Texas Women's U.	−0.30**	(0.03)	−0.21**	(0.03)	−0.12**	(0.03)	−0.08**	(0.03)	0.00	(0.05)
Texas A&M U.	–		–		–		–		–	
Texas Tech U.	−0.18**	(0.01)	−0.18**	(0.01)	−0.07**	(0.01)	−0.05**	(0.01)	0.00	(0.03)
S. F. Austin State U.	−0.28**	(0.01)	−0.25**	(0.01)	−0.16**	(0.01)	−0.12**	(0.01)	−0.01	(0.03)
Tarleton State U.	−0.28**	(0.02)	−0.29**	(0.02)	−0.14**	(0.02)	−0.10**	(0.02)	−0.01	(0.05)
UT Dallas	−0.12**	(0.02)	−0.12**	(0.02)	−0.14**	(0.02)	−0.11**	(0.02)	−0.02	(0.05)
West Texas A&M U.	−0.33**	(0.02)	−0.32**	(0.02)	−0.13**	(0.03)	−0.09**	(0.03)	−0.03	(0.06)
U. of North Texas	−0.31**	(0.01)	−0.28**	(0.01)	−0.20**	(0.01)	−0.16**	(0.01)	−0.04	(0.03)
UT Pan American	−0.52**	(0.01)	−0.47**	(0.02)	−0.19**	(0.03)	−0.14**	(0.03)	−0.04	(0.03)
TAMU Galveston	−0.21**	(0.04)	−0.21**	(0.04)	−0.17**	(0.04)	−0.13**	(0.04)	−0.04	(0.07)
TAMU Commerce	−0.36**	(0.02)	−0.33**	(0.02)	−0.25**	(0.03)	−0.20**	(0.03)	−0.05	(0.05)
Texas State U.	−0.24**	(0.01)	−0.23**	(0.01)	−0.13**	(0.01)	−0.10**	(0.01)	−0.05	(0.03)
UT Austin	−0.13**	(0.01)	−0.12**	(0.01)	−0.12**	(0.01)	−0.11**	(0.01)	−0.06**	(0.02)
TAMU Corpus Christi	−0.31**	(0.02)	−0.29**	(0.02)	−0.17**	(0.02)	−0.13**	(0.02)	−0.07	(0.04)
S. Houston State U.	−0.29**	(0.02)	−0.25**	(0.02)	−0.18**	(0.02)	−0.14**	(0.02)	−0.07*	(0.03)
Lamar U.	−0.35**	(0.02)	−0.32**	(0.02)	−0.21**	(0.03)	−0.16**	(0.03)	−0.08	(0.04)
UT Arlington	−0.25**	(0.02)	−0.22**	(0.02)	−0.17**	(0.02)	−0.14**	(0.02)	−0.09*	(0.04)
UT Permian Basin	−0.30**	(0.04)	−0.28**	(0.04)	−0.13*	(0.05)	−0.08	(0.05)	−0.11	(0.12)
U. of Houston	−0.26**	(0.01)	−0.21**	(0.01)	−0.20**	(0.01)	−0.17**	(0.01)	−0.11**	(0.03)
Prairie View A&M U.	−0.59**	(0.02)	−0.41**	(0.02)	−0.28**	(0.02)	−0.23**	(0.02)	−0.12**	(0.04)
TAMU Kingsville	−0.50**	(0.02)	−0.46**	(0.02)	−0.23**	(0.03)	−0.18**	(0.03)	−0.12**	(0.04)
Midwestern State U.	−0.42**	(0.02)	−0.40**	(0.02)	−0.23**	(0.03)	−0.18**	(0.03)	−0.13**	(0.05)
Angelo State U.	−0.43**	(0.02)	−0.41**	(0.02)	−0.20**	(0.02)	−0.16**	(0.02)	−0.14**	(0.04)
UT San Antonio	−0.40**	(0.01)	−0.37**	(0.01)	−0.24**	(0.02)	−0.19**	(0.02)	−0.14**	(0.03)
U. of Houston Dwtn.	−0.42**	(0.02)	−0.33**	(0.02)	−0.22**	(0.02)	−0.17**	(0.02)	−0.14**	(0.04)
Sul Ross State U.	−0.48**	(0.04)	−0.45**	(0.04)	−0.19**	(0.04)	−0.14**	(0.04)	−0.15	(0.08)
TAMU International	−0.41**	(0.03)	−0.36**	(0.03)	−0.26**	(0.05)	−0.20**	(0.05)	−0.18**	(0.06)
Texas Southern U.	−0.78**	(0.02)	−0.60**	(0.02)	−0.43**	(0.03)	−0.37**	(0.03)	−0.23**	(0.05)
UT El Paso	−0.63**	(0.02)	−0.58**	(0.02)	−0.40**	(0.03)	−0.35**	(0.03)	−0.27**	(0.05)
Year FE	Yes		Yes		Yes		Yes		Yes	
Race FE, gender			Yes		Yes		Yes		Yes	
H.S. FE, H.S. demographics					Yes		Yes		Yes	
SAT score, SAT demographics							Yes		Yes	
Application group FE									Yes	
Observations	169,239		169,239		169,239		169,239		169,239	
Adjusted R^2	0.03		0.03		0.05		0.06		0.06	

Notes: (1) Sample includes all graduates of Texas public high schools between 1998 and 2002 who ever enrolled in a Texas public college and had non-zero earnings reported in their state unemployment insurance record eight calendar years after graduating from high school. (2) The excluded school is Texas A&M University. TAMU, Texas A&M University; UT, University of Texas. (3) High school demographics include indicators for enrollment in various courses and programs (see text), eligibility for free lunch, and a school-defined indicator of being at risk of not graduating. (4) SAT demographics include survey questions collected by the College Board on a student's subjective rank in high school, high school GPA, father's and mother's education level, household income, and planned educational attainment.

* $p < 0.05$.

** $p < 0.01$.

unconditional earnings of students who initially enroll at Texas A&M are significantly higher than those of students who initially enroll in any other four-year public institution in Texas. While this is useful information, we are ultimately interested in all pairwise tests of significance between colleges. We return to discuss this issue below.

The results in column 1 serve as a base case that can be compared with models that control for selection into colleges. Columns 2 through 5 sequentially add the following sets of control variables: (i) race and gender controls, (ii) high school fixed effects and indicators for courses taken during high school, (iii) SAT score and various student and family demographics collected by the College Board, and (iv) application group fixed effects. The order of addition roughly reflects the likely availability of data in administrative databases: race and gender are likely available in any higher education database; high school information requires coordination with the state K–12 education agency; SAT information must be purchased from the College Board; college application data is increasingly common, but costly to collect.

Our main interest is in how estimates of β_s change upon controlling for various covariates. Comparing across the columns of Table 3, several observations are worth noting. First, the range of point estimates shrinks as more controls are added. For example, the absolute range is 0.78 in column 1, 0.43 in column 3 and 0.27 in column 5. This trend could be explained by students selecting into colleges according to innate ability (which leads to higher earnings); for example, students with high innate ability are more likely to attend Texas A&M, while students with relatively lower innate ability are more likely to attend Texas Southern University. Therefore, controlling for these predetermined covariates that are correlated with ability reduces the disparity in earnings across institutions perceived to be of high or low quality, bringing the estimates closer to the true causal effects of attending particular colleges.

Second, some sets of covariates change resulting estimates much more than others. For example, moving from column 1 to column 2 – controlling for race and gender in addition to year fixed effects – changes coefficients much less than moving from, say, column 2 to column 3 – additionally controlling for high school fixed effects and high school course-taking indicators.

Third, the model with the most comprehensive set of covariates currently available in Texas stands out because many of the differences between Texas A&M and other universities become insignificant, as seen in column 5. This difference appears despite the fact that the application group fixed effects only add information for the 37 percent of the sample that applied to more than one school (see Table 2). The estimates in column 5 imply we cannot statistically distinguish differential value added across schools in this model. While we do not believe that the model in column 5 has fully solved the omitted variable-bias problem, it is likely that controlling for even more covariates that determine the selection process – such as the quality of one's entrance essays or letters of recommendation – will reduce significant differences even further.[12]

5.2. Significant differences across colleges

While informative, Table 3 does not directly inform us of the difference in earnings between all pairwise combinations of institutions. One way to summarize this information is demonstrated in Table 4, which contains a matrix of *p*-values from pairwise tests of significance of the coefficients in column 5 of Table 3, our preferred estimates that match students on their college application and acceptance profiles.

The columns and rows in Table 4 are ordered as in Table 3, by order of the highest conditional earnings from the fully specified model, and these *p*-values are adjusted using the Benjamini–Hochberg methodology (Benjamini & Hochberg, 1995) to reflect the fact that multiple hypotheses are being performed, an important consideration considering that the probability of rejecting a true null hypothesis is greater than the significance level used for each test. Comparisons between all colleges are now visible and, for example, we cannot say with any degree of confidence that students attending UT Permian Basin, a small regional college located in West Texas, earn less on average than observationally equivalent students at any other college. However, this is due to a large standard error that is partially a result of the relatively small number of students attending UT Permian Basin.

5.3. Outcomes: persistence and graduation rates

We now investigate one-year persistence rates and graduation rates as outcomes. In the interest of brevity, we display in Table 5 only results from the model that includes the largest set of covariates used in column 5 of Table 3; however, qualitatively similar patterns as in Table 3 result from the successive addition of covariates with persistence and graduation as outcomes. Columns 1 and 3 of Table 5 include all college enrollees, while columns 2 and 4 contain only enrollees with non-zero annual earnings (the same sample as used in the preceding earnings models). For ease of comparison, we order colleges in Table 5 as in Table 3.

5.3.1. Comparing across outcome measures

Focusing first on columns 1 and 3, it is obvious that the ranking of schools is different from the ranking in the earnings regressions. For example, UT Tyler does relatively poorly in conditional graduation rates (18 percentage points lower than Texas A&M), although it had the highest conditional earnings in the state. On the other hand, UT El Paso, for example, has the lowest conditional earnings and persistence rate amongst all public institutions in Texas.

It is important to note that different outcomes (e.g., earnings versus graduation rates) capture different parts of the education production process and may lead to different value-added rankings. Nevertheless, there is certainly a correlation between persistence and graduation, in that one must persist to graduate, and there is likely a correlation between graduation and earnings, in that years in college have a positive return in the labor market, at least marginally. This correlation can be seen graphically in Fig. 1, which contains pairwise scatter plots of the coefficients on the school fixed effects from the earnings, graduation, and persistence models, and in Panel A of Table 6, which contains the matrix of correlation coefficients for the estimates from these three models. The correlation is strongly positive between all the measures, with the strongest association between the graduation and persistence models.

It may also be of interest to think about the ranking of schools in terms of their differential ability to add value to their students. In this case, the rank correlation between the different outcomes is of interest. Indeed, as seen in Panel B of Table 6, the magnitude of the rank

[12] The models in Table 3 pool together all cohorts in our data, which increases the precision of value-added estimates, but comes at a cost of aggregating the effects of any administrative policies that vary over time. This tradeoff is important for policymakers interested in using value-added measures to incentivize short-term behavior. We show in Online Appendix Table 2 that there is a moderate amount of movement in value-added estimates across cohorts in our data. As our model cannot isolate whether this volatility reflects changes in administrative policy or natural variations in cohorts over time, this is further reason to think of institutional performance in terms of broad classifications of institutional performance, as opposed to specific ordinal rankings.

Table 4
p-values from tests of significant differences between coefficients on college enrollment indicators in the fully specified earnings model.

	UT Tyler	Texas Women's U.	Texas A&M U.	Texas Tech U.	S. F. Austin State U.	Tarleton State U.	UT Dallas	West Texas A&M U.	U. of North Texas	UT Pan American	TAMU Galveston	TAMU Commerce	Texas State U.	UT Austin	TAMU Corpus Christi	S. Houston State U.	Lamar U.	UT Arlington	UT Permian Basin	U. of Houston	Prairie View A&M U.	TAMU Kingsville	Midwestern State U.	Angelo State U.	UT San Antonio	U. of Houston Dwtn.	Sul Ross State U.	TAMU International	Texas Southern U.	UT El Paso
UT Tyler		.99		.99	.91	.94	.91	.86	.76	.78	.81	.78	.71	.64	.60	.60	.55	.49	.66	.39	.32	.33	.32	.25	.24	.28	.39	.21	.05	.03
Texas Women's U.				.99	.87	.92	.89	.83	.60	.69	.78	.68	.54	.43	.43	.41	.39	.27	.63	.17	.11	.17	.16	.09	.05	.11	.29	.10	.00	.00
Texas A&M U.																														
Texas Tech U.					.78	.91	.85	.78	.37	.51	.74	.57	.24	.11	.24	.20	.21	.09	.60	.01	.01	.05	.05	.01	.00	.03	.24	.04	.00	.00
S. F. Austin State U.						.95	.99	.91	.64	.76	.85	.75	.51	.33	.41	.32	.35	.22	.68	.06	.04	.11	.12	.04	.00	.06	.32	.07	.00	.00
Tarleton State U.							.95	.90	.71	.78	.83	.77	.64	.52	.51	.49	.46	.35	.67	.21	.16	.20	.20	.11	.06	.15	.33	.12	.01	.00
UT Dallas								.92	.78	.83	.88	.81	.71	.58	.57	.56	.51	.39	.70	.24	.20	.24	.23	.14	.09	.20	.37	.14	.01	.00
West Texas A&M U.									.92	.95	.95	.92	.90	.81	.78	.78	.71	.64	.77	.49	.41	.43	.40	.29	.28	.35	.49	.26	.05	.03
U. of North Texas										.95	.99	.97	.92	.78	.74	.72	.67	.48	.78	.24	.19	.27	.26	.14	.04	.20	.45	.17	.00	.00
UT Pan American											.97	.95	.89	.74	.69	.69	.64	.49	.78	.26	.20	.21	.26	.14	.04	.20	.43	.15	.00	.00
TAMU Galveston												.98	.95	.91	.85	.87	.81	.77	.81	.63	.54	.54	.51	.42	.41	.46	.56	.33	.11	.06
TAMU Commerce													.97	.90	.83	.83	.78	.69	.81	.49	.39	.43	.41	.29	.24	.33	.51	.25	.03	.02
Texas State U.														.84	.78	.78	.71	.55	.81	.27	.20	.29	.29	.15	.04	.20	.49	.18	.00	.00
UT Austin															.91	.91	.83	.69	.85	.35	.27	.37	.37	.20	.06	.26	.54	.21	.01	.00
TAMU Corpus Christi																.98	.94	.88	.90	.68	.54	.54	.54	.40	.29	.46	.64	.32	.04	.02
S. Houston State U.																	.91	.83	.89	.58	.43	.51	.49	.33	.23	.39	.61	.29	.02	.01
Lamar U.																		.94	.91	.78	.64	.67	.64	.49	.43	.54	.69	.39	.06	.04
UT Arlington																			.94	.83	.69	.71	.67	.51	.43	.58	.71	.41	.06	.03
UT Permian Basin																				.99	.95	.95	.93	.90	.91	.91	.90	.79	.58	.45
U. of Houston																					.85	.85	.81	.66	.60	.71	.81	.51	.06	.05
Prairie View A&M U.																						.97	.92	.81	.83	.87	.89	.64	.12	.09
TAMU Kingsville																							.95	.88	.90	.91	.91	.69	.24	.14
Midwestern State U.																								.92	.95	.95	.93	.77	.32	.20
Angelo State U.																									.95	.95	.98	.81	.35	.20
UT San Antonio																										.99	.95	.75	.21	.11
U. of Houston Dwtn.																											.95	.78	.29	.20
Sul Ross State U.																												.90	.63	.44
TAMU International																													.70	.49
Texas Southern U.																														.78
UT El Paso																														

Note: Cells contain Benjamini–Hochberg adjusted *p*-values from pairwise tests of significant differences between coefficients on college enrollment indicators in the fully specified earnings model. This model controls for year fixed effects, student demographics from both high schools and the College Board, high school fixed effects, SAT scores, and application group fixed effects.

Table 5
Graduation and persistence differences of enrollees at Texas four-year public colleges. Conditioning on the full set of observable covariates.

Sample =	All enrollees		Enrollees with non-zero earnings		All enrollees		Enrollees with non-zero earnings	
Outcome =	Graduated		Graduated		Persisted to year 2		Persisted to year 2	
	(1)		(2)		(3)		(4)	
UT Tyler	−0.18**	(0.04)	−0.15**	(0.04)	−0.08**	(0.03)	−0.06	(0.03)
Texas Women's U.	−0.10**	(0.02)	−0.07**	(0.02)	−0.02	(0.02)	−0.01	(0.02)
Texas A&M U.	–		–		–		–	
Texas Tech U.	−0.03**	(0.01)	−0.02*	(0.01)	0.01*	(0.01)	0.01	(0.01)
S. F. Austin State U.	−0.10**	(0.01)	−0.08**	(0.01)	−0.08**	(0.01)	−0.07**	(0.01)
Tarleton State U.	−0.08**	(0.02)	−0.08**	(0.02)	−0.04**	(0.01)	−0.04*	(0.02)
UT Dallas	−0.06**	(0.02)	−0.06**	(0.02)	−0.01	(0.01)	−0.01	(0.01)
West Texas A&M U.	−0.14**	(0.02)	−0.12**	(0.03)	−0.07**	(0.02)	−0.05*	(0.02)
U. of North Texas	−0.10**	(0.01)	−0.09**	(0.01)	−0.02**	(0.01)	−0.01	(0.01)
UT Pan American	−0.11**	(0.01)	−0.10**	(0.01)	−0.06**	(0.01)	−0.05**	(0.01)
TAMU Galveston	−0.03	(0.02)	−0.02	(0.03)	−0.02	(0.02)	−0.00	(0.02)
TAMU Commerce	−0.12**	(0.02)	−0.09**	(0.02)	−0.03*	(0.02)	−0.01	(0.02)
Texas State U.	−0.04**	(0.01)	−0.04**	(0.01)	0.01	(0.01)	0.01	(0.01)
UT Austin	−0.04**	(0.01)	−0.04**	(0.01)	0.00	(0.00)	0.00	(0.00)
TAMU Corpus Christi	−0.07**	(0.02)	−0.05**	(0.02)	−0.01	(0.01)	−0.00	(0.01)
S. Houston State U.	−0.06**	(0.01)	−0.05**	(0.01)	−0.05**	(0.01)	−0.04**	(0.01)
Lamar U.	−0.20**	(0.01)	−0.18**	(0.02)	−0.11**	(0.01)	−0.08**	(0.01)
UT Arlington	−0.13**	(0.01)	−0.12**	(0.01)	−0.06**	(0.01)	−0.06**	(0.01)
UT Permian Basin	−0.10*	(0.05)	−0.10*	(0.05)	0.03	(0.03)	0.05	(0.03)
U. of Houston	−0.15**	(0.01)	−0.13**	(0.01)	−0.03**	(0.01)	−0.03**	(0.01)
Prairie View A&M U.	−0.08**	(0.01)	−0.07**	(0.01)	−0.01	(0.01)	−0.01	(0.01)
TAMU Kingsville	−0.11**	(0.02)	−0.09**	(0.02)	−0.05**	(0.01)	−0.05**	(0.02)
Midwestern State U.	−0.14**	(0.02)	−0.12**	(0.02)	−0.05**	(0.01)	−0.05**	(0.01)
Angelo State U.	−0.15**	(0.01)	−0.14**	(0.02)	−0.08**	(0.01)	−0.06**	(0.01)
UT San Antonio	−0.18**	(0.01)	−0.17**	(0.01)	−0.08**	(0.01)	−0.07**	(0.01)
U. of Houston Dwtn.	−0.29**	(0.02)	−0.28**	(0.02)	−0.06**	(0.02)	−0.04*	(0.02)
Sul Ross State U.	−0.17**	(0.03)	−0.15**	(0.03)	−0.09**	(0.02)	−0.07*	(0.03)
TAMU International	−0.13**	(0.02)	−0.11**	(0.03)	−0.12**	(0.02)	−0.10**	(0.02)
Texas Southern U.	−0.25**	(0.01)	−0.24**	(0.02)	−0.05**	(0.01)	−0.05**	(0.01)
UT El Paso	−0.22**	(0.02)	−0.21**	(0.02)	−0.12**	(0.02)	−0.11**	(0.02)
Year FE, H.S. FE, application group FE, SAT score, student demographics	Yes		Yes		Yes		Yes	
Observations	217,733		169,239		217,733		169,239	
Adjusted R^2	0.20		0.19		0.10		0.10	

Notes: (1) "All enrollees" sample includes all graduates of Texas public high schools between 1998 and 2002 who ever enrolled in a Texas public college. "Enrollees with non-zero earnings" restricts the "All enrollees" sample to those who had strictly positive earnings reported in state Unemployment Insurance records eight years after graduating from high school. (2) Graduating is defined as the student receiving a bachelor's degree at any public or private four-year college in Texas by the end of the sixth academic year after graduating from high school. Persistence into Year 2 of college is defined as the student completing at least 30 credit hours at any public or private four-year college in Texas by the end of the sixth academic year after graduating high school. (3) The excluded school is Texas A&M University. TAMU, Texas A&M University; UT, University of Texas. (4) Student demographics include high school variables (indicators for enrollment in various courses and programs (see text), eligibility for free lunch, a school-defined indicator of being at risk of not graduating), and SAT variables (survey questions collected by the College Board on a student's subjective rank in high school, high school GPA, father's and mother's education level, household income, and planned educational attainment.

* $p < 0.05$.
** $p < 0.01$.

correlation is again strongly positive between all outcome measures.

5.3.2. Sensitivity to the choice of sample

The practical use of this value-added model requires several sample selection criteria; to name a few, whether to include or exclude out-of-state enrollees in in-state colleges, students in graduate school when earnings is the outcome, or students with missing pre-enrollment covariates. Importantly, such sample selection criteria may influence resulting value-added estimates.

We demonstrate the sensitivity of the choice of sample with one example salient to our empirical context: whether or not to include students with zero earnings when using graduation and persistence as outcomes. There is no a priori reason why we would want to exclude enrollees with missing earnings when considering graduation or persistence, but if we are comparing estimates from persistence and graduation rate models to earnings models, it is useful to use the same sample. This sensitivity can be seen by comparing column 1 with column 2, and column 3 with column 4 in Table 5. In this case, despite excluding about 22 percent of the sample, there is little movement in the point estimates across the samples. However, the general point remains that different samples can lead to different value-added estimates.

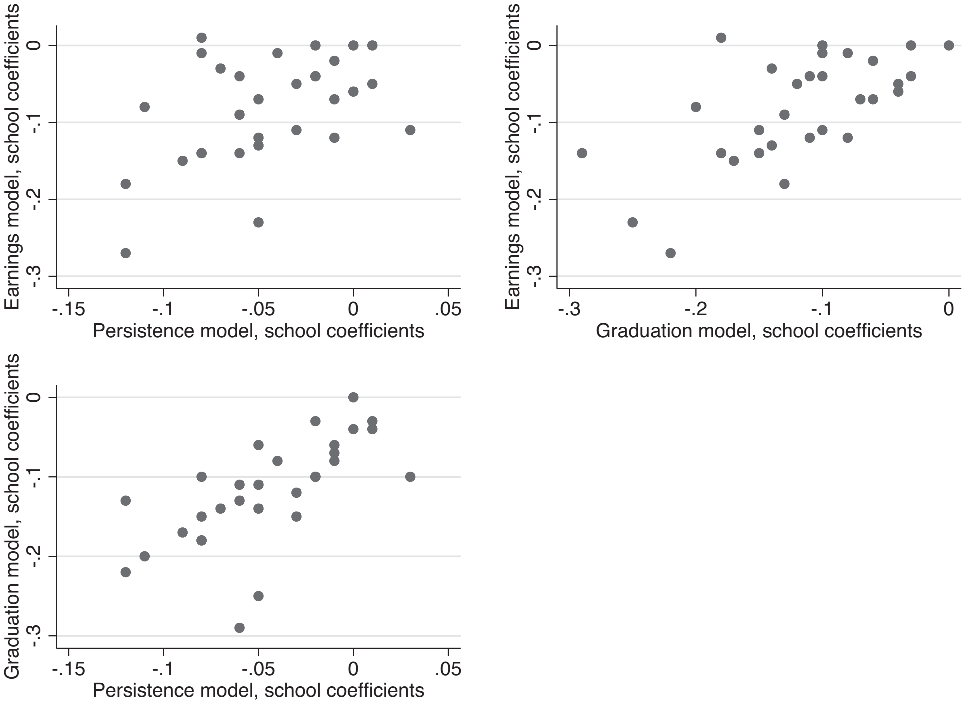

Fig. 1. Scatter plots of the coefficients on school fixed effects in the earnings, graduation, and persistence models.

Table 6
Correlations between the school coefficients (Panel A) and the rank-order of school coefficients (Panel B) from the earnings, graduation, and persistence models.

Panel A			
	Earnings model, coefficient	Graduation model, coefficient	Persistence model, coefficient
Earnings model, coefficient	1.000		
Graduation model, coefficient	0.640	1.000	
Persistence model, coefficient	0.467	0.666	1.000
Panel B			
	Earnings model, rank	Graduation model, rank	Persistence model, rank
Earnings model, rank	1.000		
Graduation model, rank	0.624	1.000	
Persistence model, rank	0.455	0.800	1.000

6. Conclusions and recommendations

In this paper, we have discussed the issues surrounding the measurement of value added in higher education, and we proposed a general methodology for measuring the value added of individual institutions using existing administrative data. Our empirical exercise estimates value-added models for the state of Texas considering persistence rates, graduation rates, and earnings as outcome measures. We demonstrate how the choice of both conditioning variables and the base sample can influence value-added estimates, and we emphasize the importance of statistically distinguishing value-added estimates between individual institutions.

Owing to the multidimensional objective function of higher education institutions, it is unlikely that a single outcome measure will fully reflect institutional performance. Thus, we have argued that a system of multiple metrics capturing various dimensions of institutional performance is crucial to gain a full understanding of overall institutional performance. We believe further research should consider the usefulness of optimally combining multiple measures of college performance, perhaps along the lines of recent work in the health care sector (e.g., McClellan & Staiger, 2000).

Regardless of the outcome in question, the interest in value-added measures in part stems from a desire amongst policymakers to incentivize performance, as is evidenced by the current debate in many U.S. states. Importantly, our results suggest that policymakers should use caution in applying value-added methods of the kind we estimate for incentivizing purposes or otherwise tie them to pay or funding. We show that value-added measures change significantly upon conditioning on successively more information about the pre-enrollment characteristics of students, and we argue that even conditioning on the most complete set of characteristics available is unlikely to yield value-added estimates that can be interpreted causally. As such, improperly designed measures have as much potential to deceive as to inform about true institutional quality.

We have also shown that adjusting outcomes for predetermined student characteristics often makes it impossible to statistically distinguish across many sets of colleges. As such, we urge policymakers to think in terms of broad classifications of institutional performance, as opposed to specific ordinal rankings. One possibility is to determine a benchmark (e.g., a target graduation rate) and classify institutions relative to that benchmark (e.g., statistically below, within normal statistical variance of, or exceeding the specified benchmark).

Finally, experience with value-added measures in the K–12 sector has shown that institutions are likely to game the system (e.g., Figlio, 2006), and there is no reason to not expect similar responses in higher education. For example, if funding is tied to conditional completion rates, institutions may lower grading standards in order to graduate more students or admit students with unobserved attributes that make them more likely to graduate. Policy must acknowledge and address such perverse incentives when developing and implementing value-added measures for higher education.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.econedurev.2014.06.001.

References

Acemoglu, D., & Angrist, J. (2001]). How large are human-capital externalities? Evidence from compulsory-schooling laws In *NBER macroeconomics annual 2000, Vol. 15, NBER Chapters* (pp. 9–74). Cambridge, MA: National Bureau of Economic Research, Inc..
Andrews, R. J., Li, J., & Lovenheim, M. F. (2012]). *Quantile treatment effects of college quality on earnings: Evidence from administrative data in Texas. Working Paper 18068*. National Bureau of Economic Research.
Arcidiacono, P., Cooley, J., & Hussey, A. (2008]). The economic returns to an MBA. *International Economic Review, 49*(3), 873–899.
Behrman, J. R., Rosenzweig, M. R., & Taubman, P. (1996]). College choice and wages: Estimates using data on female twins. *The Review of Economics and Statistics, 78*(4), 672–685.
Benjamini, Y., & Hochberg, Y. (1995]). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological), 57*(1), 289–300.
Black, D. A., & Smith, J. A. (2004]). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics, 121*(1–2), 99–124.
Brewer, D. J., Eide, E. R., & Ehrenberg, R. G. (1999]). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. *The Journal of Human Resources, 34*(1), 104–123.
Broecke, S. (2012]). University selectivity and earnings: Evidence from UK data on applications and admissions to university. *Economics of Education Review, 31*(3), 96–107.
Cellini, S. R., & Chaudhary, L. (2012]). *The labor market returns to a for-profit college education. NBER Working Paper 18343*.
Dale, S. B., & Krueger, A. B. (2002]). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics, 117*(4), 1491–1527.
Dee, T. S. (2004]). Are there civic returns to education? *Journal of Public Economics, 88*(9–10), 1697–1720.
Dougherty, K. J., Natow, R. S., & Vega, B. E. (2012]). Popular but unstable: Explaining why state performance funding systems in the United States often do not persist. *Teachers College Record, 114*(3), 1–42.
Figlio, D. N. (2006]). Testing, crime and punishment. *Journal of Public Economics, 90*(4–5), 837–851.
Harnisch, T. L. (2011]). *Performance-based funding: A re-emerging strategy in public higher education financing*. American Association of State Colleges and Universities, Policy Matters.
Hoekstra, M. (2009]). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *Review of Economics and Statistics, 91*(4), 717–724.
Johnes, G. (2012]). Performance indicators in higher education: A survey of recent work. *Oxford Review of Economic Policy, 8*(2), 19–34.
Jongbloed, B., & Vossensteyn, H. (2001]). Keeping up performances: An international survey of performance-based funding in higher education. *cy and Management, 23*(2), 127–145.
Kain, T. J., & Staiger, D. O. (2008]). *Estimating teacher impacts on student achievement: An experimental evaluation. NBER Working Paper 14607*.
Klein, S. P., Kuh, G., Chun, M., Hamilton, L., & Shavelson, R. (2005]). An approach to measuring cognitive outcomes across higher education institutions. *Research in Higher Education, 46*, 251–276.
Liu, O. (2011]). Value-added assessment in higher education: A comparison of two methods. *Higher Education, 61*, 445–461.
Mallier, T., & Rodgers, T. (1995]). Measuring value added in higher education: A proposal. *Education Economics, 3*(2), 119–132.
McClellan, M., & Staiger, D. (2000]). Comparing the quality of health care providers. *Forum for Health Economics and Policy, 3*(1).
Midwestern Higher Education Compact (2009). *Completion-based Funding for Higher Education*. Technical Report..
Moretti, E. (2004]). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics, 121*, 175–212.
Oreopoulos, P., & Salvanes, K. G. (2011]). Priceless: The nonpecuniary benefits of schooling. *The Journal of Economic Perspectives, 25*(1), 159–184.
Pascarella, E. T., & Terenzini, P. T. (2005]). *How college affects students: A third decade of research*. San Francisco, CA: Jossey-Bass.
Rodgers, T. (2007]). Measuring value added in higher education: A proposed methodology for developing a performance indicator based on the economic value added to graduates. *Education Economics, 15*(1), 55–74.
Saavedra, A. R., & Saavedra, J. E. (2011]). Do colleges cultivate critical thinking, problem solving, writing and interpersonal skills? *Economics of Education Review, 30*(6), 1516–1526.
Saavedra, J. E. (2009]). *The learning and early labor market effects of college quality: A regression discontinuity analysis*. RAND Corporation mimeo.
Snyder, T., Dillow, S., & Hoffman, C. (2007]). *Digest of education statistics. National Center for Education Statistics Report*.
Spellings, M. (2006]). *A test of leadership: Charting the future of U.S. higher education. U.S. Department of Education Report*.
Stevens, D. (2002]). *Employment that is not covered by state unemployment. LEHD Technical Working Paper (TP 2002-16)*.
Tiebout, C. M. (1956]). A pure theory of local expenditures. *The Journal of Political Economy, 64*(5), 416–424.
Wright, D. L., Fox, W. F., Murray, M. N., Carruthers, C. K., & Thrall, G. (2012]). *College participation, persistence, graduation and labor market outcomes*.
Zhang, L. (2009]). A value-added estimate of higher education quality of US states. *Education Economics, 17*(4), 469–489.