CAN "VALUE ADDED" METHODS IMPROVE THE MEASUREMENT OF COLLEGE PERFORMANCE? EMPIRICAL ANALYSES AND POLICY IMPLICATIONS

Robert Kelchen
Department of Educational Policy Studies
University of Wisconsin at Madison

Douglas N. Harris
Associate Professor of Educational Policy and Public Affairs
University of Wisconsin at Madison

September 2012

HCM STRATEGISTS

ABSTRACT

Good information is critical for parents and students as they choose colleges and, increasingly, for policymakers in their decisions about funding public institutions and holding them accountable for performance. Unfortunately, we show that the information available today, such as college rankings, is of little use for any of these important decisions; rather, current information rewards colleges for attracting stronger students and for spending more money. We address these problems by taking a "value added" approach that adjusts for differences in student backgrounds and rewards colleges for spending more only when they also generate better outcomes. We use data from more than 1,200 institutions nationwide to estimate value added to one important outcome: college graduation. We also show how the value added method that is relevant to students and parents differs from the method relevant to policymakers. To the extent that the goal is to measure how effectively colleges use their resources to help students progress, the information used today appears quite misleading.

Acknowledgements

We thank Richard Arum, Sandy Baum, Charles Clotfelter, Kristin Conklin, Jesse Cuhna, Nate Johnson and Sarah Turner for their very useful comments.

INTRODUCTION

Accountability in U.S. higher education follows a market model. In contrast to K-12 education, college students choose where to apply, then "vote with their feet" and choose which college to attend among those that accept them; federal (and some state) financial aid follows the student to the chosen institution.[1] These decisions are based on student ability and preferences, market conditions, wealth constraints and one other factor that basic economic theory suggests is a key component of any well-functioning market: information. For some students, their parents, friends and relatives may provide information about their own alma maters and those of their friends and coworkers. "I have a friend who went there" is a common refrain. Other students receive information from more formal sources, such as high school guidance counselors, college websites and college admission personnel. These formal and informal sources are also interconnected with, and informed by, federally required information about graduation rates and college guides and rankings such as *U.S. News*, *Barron's* and *Washington Monthly*.

Unfortunately, the formal sources of information bear little resemblance to anything we might reasonably call "college performance." Graduation rates by themselves tell us more about who goes to a college than about how effectively the college operates to help students. Popular college rankings suffer from the same problem, failing to take into account important characteristics of students that significantly affect their likelihood of college success and are primarily outside the control of the colleges themselves. "Less-competitive" colleges may or may not contribute more to student outcomes than "competitive" ones, but that does not stop popular rating systems from branding less-competitive colleges as less desirable. The main thing we learn from these ratings is that competitive colleges charge more, spend more, attract students with stronger academic backgrounds and build stronger reputations. Reputation, in turn, is largely driven by historical circumstances and other factors that are already part of the rankings, such as students' academic background. This is circular reasoning or, at best, self-reinforcing misinformation. The result is that a higher percentage of students attending more-competitive colleges eventually graduate, but for reasons that have nothing to do with instructional quality and program performance. We do not know how well colleges use their resources to improve outcomes for the students they serve—a gap we aim to help fill with this paper.

It is not only parents and students but also policymakers who have an interest in this information problem. To help the United States remain economically competitive with other nations, policymakers have set goals for significantly increasing the number of college graduates within the next decade.[2][3] For decades, completion rates among students enrolled at four-year institutions have stagnated around 60 percent.[4] At the same time, cost per degree is rising much faster than inflation, and research suggests that many of the instructional and noninstructional programs in wide use in higher education are reducing rather than increasing productivity in terms of degree completion.[5] The rising cost of college has been attributed to

factors other than the colleges and their performance.[6] The large social and economic returns to education notwithstanding, the fact that more and more resources are required to generate similar educational outcomes gives policymakers, parents and students good reasons to be concerned about what they are getting from—and giving to—colleges and universities.[7]

Some states and the federal government are considering addressing these problems with more rigorous accountability systems in higher education. The history of state-level accountability in public higher education dates back at least to 1976, when Tennessee required colleges and universities to report graduation rates and other measures in order to be eligible for additional performance-based funding.[8] During the 1990s, a majority of states adopted small performance-based funding systems that were not large enough to change institutions' practices.[9] Although the federal government has not yet directly implemented an accountability system in higher education, the Spellings Commission report calls for such a system to be created.[10]

Neither government-based nor market-based accountability can work without accurate performance measures. This same problem arises in K-12 education as well as health care and, in response, policymakers have increasingly turned to a new approach to providing information about the performance of public institutions. Rather than simply reporting student outcomes, "value added" or "input-adjusted" measures explicitly adjust outcomes to account for differences in the students who attend schools and colleges. To obtain value added estimates in this study, we use publicly available data from the Integrated Postsecondary Education Data System (IPEDS) and The Institute for College Access and Success (College InSight) to estimate the predicted six-year graduation rates of first-time, full-time students at 1,279 four-year colleges and universities and then compare the predicted rates with the actual rates. This addresses the first problem with college rankings and other measures of college performance—that outcomes mainly reflect what the students bring to the table rather than the "value" colleges add.

The costs involved with generating these outcomes are as important as the outcomes themselves. We therefore divide the value added to graduation by the resources required. Two different resource measures are relevant: Parents and students care about how much a college contributes to the probability of graduation, *given the tuition they have to pay*. Alternatively, policymakers are primarily concerned with the costs to society, whether paid for by government funding, tuition, endowments or other sources. We estimate both types of cost-adjusted value added measures because the right performance measure is, to some extent, in the eye of the beholder.

In the next section, we discuss the various sources of college information that are currently available. We then describe our data and the approach we use to estimating value added, including how and why we take a different approach to accounting for costs. After presenting our results, we discuss remaining technical issues and related policy implications of our findings.

SOURCES OF COLLEGE PERFORMANCE INFORMATION

Students, their families and policymakers have access to a wide range of sources regarding college performance. The availability of performance information has the potential to affect where students decide to go to college, but this information can often be difficult to obtain.[11] The most direct source of information comes from the institutions themselves, through viewbooks/guides, admission counselors and websites. The experiences of family and friends with a particular college also can serve as a rough measure of an institution's performance. Many stakeholders also rely on federally required information, Internet resources, and college guides and rankings. As we show below, none of these say very much about how effectively colleges transform resources into results.

Federally required information

The federal government requires any institution receiving federal funds (including financial aid such as Pell Grants) to provide certain information about student outcomes. Since 1999, institutions have been required to provide graduation rates for first-time, full-time, degree-seeking students who enrolled during the fall semester.[12] This information is then made publicly available through the Integrated Postsecondary Education Data System (IPEDS). The College Navigator portal within IPEDS provides students and their families an opportunity to view these graduation rates (as well as other institutional characteristics) for a set of colleges; this information is also available for high school guidance counselors to use. Institutions were required under the 2008 reauthorization of the Higher Education Opportunity Act to report graduation rates for subgroups of students, such as Pell Grant recipients; however, compliance with these new requirements has been uneven at best.[13]

Internet resources

In an increasingly connected world, many students will start with general Internet searches to gather information. Prospective students and their families can find a seemingly overwhelming number of websites on the topic (a Google search for "how to choose a college" returned more than 20 million results). The available information includes websites for state and federal education agencies, articles from the news media about factors to consider in applying to colleges, websites of for-profit college search agencies (with varying amounts of useful information and sales pitches for expensive services), and promotional websites for largely for-profit colleges and universities. To be fair, at least some of the websites do recognize the importance of costs and the complexity of the factors going into college decisions and emphasize that

students should choose the colleges that are right for them, based on factors such as geography and enrollment size. These factors are no doubt associated with college success, although one would have to wonder about a college that seems like a good fit but yet has a 50 percent graduation rate.

U.S. News and other college rankings

The most popular set of college rankings is published annually by *U.S. News & World Report*. Approximately 40 percent of all students in a national survey view the *U.S. News* rankings as at least "somewhat important" in their college choice process.[14] This percentage was greatest among higher-SES students and those with better high school grades. Research has shown that a small change in the rankings, especially at top-tier universities, can drastically change the composition of the student body. A lower *U.S. News* ranking can result in a lower-quality and less diverse entering class and can require more grant aid to lure academically strong students to the institution, while moving up in the rankings results in a higher-quality student body the following year.[15]

U.S. News creates its rankings using the following measures and weights: peer assessment (25 percent), six-year graduation rate (16–20 percent), first-year retention rate (4–5 percent), faculty resources (20 percent), student selectivity (15 percent), financial resources (10 percent), alumni giving rate (5 percent) and graduation rate performance (0–5 percent).[16] Faculty resources, financial resources and student selectivity collectively represent the "inputs" that colleges have to work with, while the six-year graduation rate and first-year retention represent student "outputs."[17] Thus, we can think of these rankings as the (weighted) sum of outputs plus inputs rather than the traditional economic approach: outputs *conditional on* inputs. This has the important implication of rewarding colleges directly for using more resources, regardless of whether they use them efficiently.[18]

The peer assessment component is arguably even more problematic because it involves colleges ranking one another, inviting gaming among the colleges. Clemson University offers a striking example because its president ranked almost all other schools as below average in the peer assessment, directed potential first-time freshmen in the lower two-thirds of their high school class to technical colleges for one year and strategically lowered some class sizes that counted in the rankings while raising other class sizes that do not count in *U.S. News*.[19]

The *Washington Monthly* college rankings are intended to better measure the value added by colleges than *U.S. News* by using three equally weighted outcomes that are not as strongly correlated with inputs: social mobility, research and service. Specifically, these rankings include the following measures: the percentage of students receiving Pell Grants and the predicted graduation rate (social mobility); research expenditures, the number of students earning Ph.D.'s, faculty receiving significant awards and faculty elected to national

academies (research); and the percentage of graduates in the Peace Corps and ROTC as well as the percentage of work-study funds spent on service (service). However, even in this case, resources still constitute approximately one-third of the information used to make the rankings.

Barron's Profiles of American Colleges, published annually since 1991, rates colleges based on their level of admissions selectivity. There are six primary levels of selectivity, ranging from noncompetitive to most competitive.[20] These ratings are based solely on the percentage of students admitted and the academic characteristics of the incoming class.

All three of these rankings reflect an underlying assumption that good colleges are those that attract academically talented and wealthy students, rather than those whose instructional and other programs are of high quality. There is some merit to this argument. From students' and parents' perspectives, having peers with higher academic abilities can be directly beneficial for academic success.[21] Fellow students also provide important social networks, and some employers restrict their recruiting to colleges with high percentages of strong students—both of which open up useful career opportunities and provide social status.[22] However, a recent article in the *New York Times* questioned whether going to a highly prestigious college is worth the cost, and some research suggests that most students do not receive additional benefits from attending a highly prestigious college.[23]

The more a ranking system focuses on student background, the more colleges will focus on attracting talented students and the less they will focus on program quality.[24] Further, the more performance rankings focus on inputs, the more incentive there is to raise tuition and use additional resources. As a drop in rankings can be financially devastating, universities are forced to compete, often in unproductive ways, for the type of student who would help them receive a better ranking.[25]

The problem is that none of these ranking systems really have anything to do with what might reasonably be defined as "performance." We argue that any reasonable definition has to go beyond measures of peer quality and prestige to include the quality of course offerings and instruction and how well noninstructional programs facilitate graduation and long-term success. This is especially true today, with growing evidence that overall college productivity is on the decline and popular and expensive programs are inefficient.[26]

College rankings and value added

Researchers have realized the problem of comparing the effectiveness of institutions with heterogeneous student bodies, and this is why there has been a shift toward value added. This approach compares actual student outcomes with those predicted based on student characteristics and other factors that are outside the control of the institution.[27] In particular, the approach accounts for what is arguably the most important

factor driving student outcomes—the students themselves. Students enter college with very different probabilities of achieving important outcomes (such as credits attained, year-to-year persistence and graduation), and the job of colleges is to increase the probabilities as much as possible. Accounting for student heterogeneity is therefore critical for measuring what colleges really contribute to student outcomes. Value added measures have rapidly expanded in K-12 education along with increased standardized testing.[28] With annual student outcome measures, it is possible to account for past performance and thereby account for most of the factors that drive future student outcomes.

One alternative way to capture instructional and program performance is to follow the lead of the *U.S. News* rankings and include "peer" ratings from leaders of other colleges, though these leaders have limited familiarity with institutions other than their own and have incentives to "game" the system to increase their own rankings. The giving rate among alumni may also capture value added to some degree, although it says little about current performance because many alumni donors attended college decades ago. Therefore, neither alumni nor peer institution ratings are likely to get us very far in improving performance measures.

Recognizing this, in 1997, *U.S. News* began to apply value added-like techniques to create a "graduation rate performance" measure that now constitutes a small portion of its rankings. Specifically, *U.S. News* predicts graduation rates using per-student academic expenditures, standardized test scores and the percentage of students who are receiving Pell Grants.[29] The difference between predicted and actual graduation rates then becomes the value added measure. There are some problems with the specifics of the *U.S News* approach to value added, as we discuss later, but the key point is that graduation rate performance constitutes only 7.5 percent of the ranking at research and liberal arts institutions and is excluded for other institutions.[30]

In 2005, *Washington Monthly* followed suit by including a graduation rate performance measure in its college rankings. The magazine estimates value added through an average of two measures. In the first measure, it regresses graduation rates on the percentage of students who receive Pell Grants and the 25th and 75th percentile scores on a standardized test (SAT or ACT).[31] In the second measure, it regresses graduation rates on the predicted percentage of students receiving Pell Grants given the standardized test score; that value is then used to predict the graduation rate.[32] Graduation rate performance makes up a relatively small portion of the ranking for all institutions.

A less well-known system, College Results Online, produced by The Education Trust, also carries out value added-like calculations.[33] The website does not yield a fixed set of comparisons but rather allows students to select a single college, and then the site identifies "similar" colleges based on an index calculated from measures like those we use to estimate value added. The graduation rates and other information from these similar colleges are then reported. Although College Results Online does not report value added estimates, it

does provide a method to compare the outcomes of reasonably similar colleges. One problem, however, is that it is very difficult for a computer to identify the comparisons each student really wants to make. Students do not always limit their searches to "similar" institutions, especially as most college guides and many guidance counselors now recommend having "reach" and "safety" colleges in the application process. Policymakers also need to know which types of colleges are most efficient to make wise decisions about funding and other higher education policies, which necessarily requires comparing the *dissimilar* colleges they oversee.

Again, peer characteristics and prestige are important, but it is hard to argue that instructional and service quality should be almost completely ignored as they are now. The rankings now are almost a self-fulfilling prophecy: Expensive colleges can charge more because they have better reputations among peers and loyal, wealthy alumni. By spending more, they improve their rankings and further improve their reputations. Comparing outcomes for "similar" institutions helps some, but similarity is itself being defined in terms of some of the same criteria that are now producing misleading characterizations of performance. We argue that measuring instructional and program effectiveness is important and that there is a better way.

DATA AND METHODS

To estimate value added based on graduation rates, we use three years of publicly available data from the Integrated Postsecondary Education Data System (IPEDS) and The Institute for College Access and Success (College InSight). Institutions of higher education that receive any federal funding are required to provide data for IPEDS, while providing data for the relevant portion of the College InSight data set is optional. The IPEDS data set includes background institutional and student characteristics through the 2008–2009 academic year, and the College InSight data set includes information about the percentage of students receiving Pell Grants through the 2008–2009 academic year.[34] We started with the 1,637 four-year institutions in IPEDS that primarily offer four-year degrees, and we eliminated all institutions that do not report a six-year graduation rate in the 2008–2009 academic year; that generally do not offer bachelor's degrees; or that do not report either (a) ACT or SAT scores or (b) the percentage of students receiving Pell Grants in at least one of the three academic years from 2006–2007 through 2008–2009.[35] This results in 1,288 institutions. Dropping the nine for-profit institutions gives us the final sample of 1,279 colleges and universities in all 50 states. By Carnegie classification, there are 253 research universities, 526 master's universities and 500 bachelor's universities in the full sample. The descriptive statistics of the data set are shown in Table 1.

Table 1: Summary Statistics of the Variables					
Variable	N	Mean	St. Dev.	Min	Max
Six-year graduation rate (%)	1279	55.1	18.0	1	98
Percent receiving Pell Grants	1275	29.1	14.8	0	93
Average ACT composite	1245	22.9	3.3	12	34
Education expenses per FTE	1259	$14,648	$11,562	$3,951	$146,192
Receiving financial aid (%)	1279	87.8	13.6	0	100
Receiving student loans (%)	1275	60.9	19.1	2	99
Male (%)	1279	43.0	12.2	0	100
Full-time (%)	1279	85.1	12.8	8.9	100
Undergraduate enrollment	1279	5,629	6,867	63	53,298
White (%)	1279	66.5	22.1	0	97.9
Black (%)	1279	12.5	19.3	0	99.5
Hispanic (%)	1279	6.2	8.8	0	92.7
Carnegie research (%)	1279	19.8	39.9	0	100
Carnegie master's (%)	1279	41.1	49.2	0	100
Mainly residential (%)	1279	75.9	42.8	0	100
Admit rate (%)	1268	65.1	18.3	8	100
Public (%)	1279	38.6	48.7	0	100
HBCU (%)	1279	4.8	21.3	0	100

Notes:

1) The summary statistics presented are from the 2008-09 academic year for all variables.

2) All variables shown in this table are from IPEDS and are not imputed. The percentage of students receiving Pell Grants in the 2006-07 and 2007-08 academic years (not shown here) is provided by the Institute for College Access & Success (College InSight).

3) The natural logarithm of endowment and education expenses per FTE and undergraduate enrollment are used in the analysis.

4) The six-year graduation rate listed is for first-time, full-time students who finish at the same institution at which they start college. We use IPEDS's revised cohort minus exclusions to generate the graduation rate.

5) We calculated the ACT composite score by averaging the 25th and 75 percentile scores that are provided in the IPEDS data. The average score is then rounded up to the nearest whole number. For schools that provide only SAT scores, we averaged the 25th and 75th percentile scores on the SAT verbal and mathematics sections separately and then added the two numbers to get the median SAT composite score. The SAT composite score is then converted to an ACT composite score using the ACT/SAT concordance guide (ACT, Inc., 2008).

6) Educational expenses refer to the sum of per-FTE instructional, academic support, and student services expenses.

7) We define mainly residential campuses as the Carnegie classifications of either primarily or highly residential.

Value-added approach

Most of our results are based on a simple "regression" analysis in which we estimate the relationship between graduation rate and each student and institutional characteristic in Table 1, while holding constant each of the other factors in the table.[36] This information is used to predict the graduation rate at each institution, and the difference between predicted and actual (what statisticians call the "residual") becomes the estimate of value added. While it might seem that we should also include factors such as class size, we exclude them because institutions make decisions about how to allocate resources (such as class size) that reflect the efficiency that we are trying to measure—and it is therefore something we explicitly do *not* want to account for in a value added model whose purpose is to measure how well institutions utilize their resources.[37] The situation is different in K-12 education, in which resource allocation decisions are often out of the hands of school leaders.

We therefore separate the estimation into two steps. In the first, we adjust the graduation rate based on the characteristics of students attending the college. We know, for example, that higher-income students are more likely to persist and graduate than students from low-income families (as measured through the percentage of students receiving Pell Grants). Specifically, we estimate the predicted graduation rate for College j ($GradRate_j$):

$$GradRate_j = \beta_{0j} + \beta_{1j}StudAcad_j + \beta_{2j}StudFin_j + \beta_{3j}StudDem_j + \beta_{4j}Inst_j + e_j \quad (1)$$

$StudAcad_j$ represents students' academic background, $StudFin_j$ represents students' financial information, $StudDem_j$ is a vector of student demographic characteristics such as race and gender, and $Inst_j$ is a vector of institutional characteristics such as Carnegie classification and type of institutional control.[38]

The residual term e_j is the difference between the predicted graduation rate and the actual graduation rate, which is our estimate of the (unadjusted) value added measure for College j. A key assumption behind this approach is that the residual term reflects only differences in college performance and not other differences between students that affect graduation that we cannot measure. For example, it seems likely that students who score 1400 on the SAT and attend Western Kentucky University are probably different from students with the same scores who attend Princeton. How different are these students? There is no way to know for sure. Even after accounting for the other factors in the equation, there are probably differences among students that we cannot measure. This means the residual likely captures part of what we do want—college performance—and differences in students that are unrelated to college performance.

Cost adjustments

The above calculations do not account for costs and therefore fall short of a true value added measure in our view. In the K-12 research on value added, costs are ignored because they often cannot be measured at the institutional level and, even when they can, resources are not within the control of K-12 school leaders in the

same way they are in colleges. Therefore, our next step is to take costs into account in our measures of college value added by dividing this value added estimate by an appropriate cost measure. In order to make meaningful comparisons across institutions, we first normalize value added so the lowest institution has an estimate of zero. Otherwise, for example, an extremely inexpensive college with a slightly negative unadjusted value added would have lower value added than one with a slightly positive unadjusted value added that uses three times the resources. Without this normalization, the negative unadjusted figures could give the false impression that many colleges are harming students—a "negative effect"—which would be quite misleading.

Policymakers are concerned about the costs to society. Therefore, this cost adjustment (CA) is made based on total educational expenses (in thousands):

$$VA_{CA}(Policy)_j = \frac{ValueAdded_j}{EdExp_j} \quad (2)$$

Educational expenses include the instructional, academic support, and student services expenses categories from IPEDS.[39] We include private institutions as well as public institutions in our models because we are starting from the standpoint of market accountability.

In the student/family model, we divide by the net cost of attendance, which is the total cost of attendance (tuition plus room and board) less the average grant aid received by students at the college (in thousands) ($NetCost_j$):

$$VA_{CA}(Family)_j = \frac{ValueAdded_j}{NetCost_j} \quad (3)$$

The numerator—value added—is identical in both models. An important implication of our approach is that some colleges with below-average value added have high cost-adjusted value added because they use so few resources. We estimate the overall and cost-adjusted value added models separately for the three academic years from 2006–07 to 2008–09, as well as using all three years of data in a panel.[40]

Rationale and possible alternative ways to incorporate costs

In many ways, our cost-adjusted value added measure is simple and intuitive. Both families and policymakers care about both outcomes and costs. Dividing the added value to graduation by costs therefore provides a natural metric for "bang for the buck." This approach is also consistent with the effectiveness-cost ratio, common in program evaluation.[41] Like cost-adjusted value added, the effectiveness-cost ratio divides the estimated impact of a program by the cost of resources involved in generating that impact. This is helpful for decisionmakers who might be choosing, for example, whether to invest more resources in class size reduction versus higher staff salaries. The same basic logic applies to policymakers who are making decisions about how to invest in public higher education. Implicitly, state legislators make decisions every year about which institutions warrant the most public resources, and they ultimately choose what they see as an optimal

mix.[42] Put differently, each college can be viewed as a collection of individual programs, making the effectiveness-cost ratio an intuitive metric for the "collective efficiency" of the institution.[43]

Family decisions about where students should attend college are a bit different, however. They are not choosing an optimal mix, but rather a single institution that they hope will meet their needs. Yet cost-adjusted value added is still a very useful metric even in this context. Like policymakers, the well-being of families and the role of higher education depend on both costs and benefits. Unadjusted value added is clearly one aspect of the benefits, and the net cost of attendance reflects what the average family has to pay for a year of college.[44] Therefore, even if the specific measures of costs might differ, the basic logic of cost-adjusted value added still applies.

FINDINGS

Unadjusted value-added calculations

In Table 2, we present the results from three different regression models for the entire sample of 1,279 institutions using data from the 2008–09 academic year.[45] In the first model, we use only the variables in the *U.S. News* graduation rate prediction (the percentage of students receiving Pell Grants, average ACT composite score and educational expenses per full-time equivalent student) as well as dummy variables for Carnegie classifications. The second model includes key student demographic variables, while the third model includes demographic variables as well as institutional factors that are difficult for a college to control.

Table 2: Regression Results--All Universities

	Model (1)--*U.S. News*		Model (2)		Model (3)	
Variable	Coeff	(SE)	Coeff	(SE)	Coeff	(SE)
Percent Pell	-.258	(.029)***	-.298	(.035)***	-.240	(.035)***
Average ACT composite	.033	(.002)***	.034	(.001)***	.028	(.001)***
Log education expenses per FTE	.031	(.012)**	--	--	--	--
Percent receiving financial aid	--	--	-.091	(.028)***	-.111	(.028)***
Percent receiving student loan	--	--	.179	(.020)***	.136	(.021)***
Percent male students	--	--	-.108	(.024)***	-.087	(.024)***
Percent full-time students	--	--	.292	(.025)***	.231	(.026)***
Log enrollment	--	--	.008	(.003)***	.025	(.005)***
Percent white students	--	--	-.053	(.026)**	.008	(.027)
Percent black students	--	--	-.024	(.032)	-.116	(.039)***
Percent Hispanic students	--	--	.074	(.047)	.081	(.045)*
Carnegie research	-.004	(.008)	--	--	.007	(.010)
Carnegie master's	.002	(.007)	--	--	.005	(.007)
Mainly residential	--	--	--	--	.035	(.008)***
Admit rate	--	--	--	--	-.091	(.018)***
Public	--	--	--	--	-.054	(.009)***
HBCU	--	--	--	--	.083	(.026)***
Constant	-.411	(.094)***	-.406	(.062)***	-.299	(.069)***
Number of colleges	1279		1279		1279	
Adjusted R-squared	.667		.717		.739	

Notes:

1) * represents p<.10, ** represents p<.05, and *** represents p<.01. Standard errors are in parentheses.

2) We used multiple imputation on a small number of observations.

Although the coefficients are generally as hypothesized, we are less focused on the role of any given factor than on the value added estimate, represented by the error term. The addition of more control variables does increase the model's explanatory power, with an adjusted R-squared of 0.739 in the full model.

One intuitive test of the validity of the unadjusted measures is whether the factors we account for are associated with graduation in the way we would expect. The vast majority of the coefficients are in the expected direction—e.g., more students receiving Pell Grants is associated with a lower graduation rate, and higher ACT/SAT scores are associated with a higher graduation rate. The main exceptions are the coefficients on race/ethnicity. While it appears that white students are less likely to graduate, this has to be interpreted relative to the groups that are omitted—particularly Asians, who are often the most likely to

succeed in college. Also, even where the signs do appear to be counterintuitive, it is important to emphasize that many of the factors affecting college success are also correlated with other factors that are measured here. This makes it harder to say whether even the most counterintuitive result really indicates a problem with the analysis.

Another possibility is that the unadjusted measures might be misleading because we implicitly assumed that the "model" of student graduation is the same at all campuses. It could be, for example, that being a racial/ethnic minority is associated with graduation at competitive colleges in a way that is different from less competitive ones. To address this, we estimated the full model from Table 2 separately for Carnegie research, master's and bachelor's institutions and then correlated the unadjusted value added measures from the separate models with those from the model for all institutions. The rank-order correlations between the two models by Carnegie classification ranged from 0.964 to 0.975. Because the rankings between the two models are very highly correlated and because the college ranking guides produce rankings roughly by Carnegie classification, we also estimate by classification in the remainder of the paper.

Our preferred approach uses three years of panel data and estimates value added separately for each Carnegie classification. These results are presented in Table 3. The results reveal some significant differences across classifications. For example, at research and master's institutions, higher rates of students living on campus and lower admit rates are significantly associated with higher predicted graduation rates; the associations are not significant at bachelor's institutions. As the coefficients are not our primary interest, we do not discuss them in detail in this paper.

Table 3: Panel Regression Results by Carnegie Classification

	Research		Master's		Bachelor's	
Variable	Coeff	(SE)	Coeff	(SE)	Coeff	(SE)
Percent Pell	-.235	(.050)***	-.213	(.034)***	-.338	(.036)***
Average ACT composite	.020	(.002)***	.015	(.002)***	.024	(.002)***
Percent receiving financial aid	-.053	(.026)**	-.061	(.028)**	-.050	(.031)
Percent receiving student loan	.009	(.023)	.063	(.019)***	.018	(.021)
Percent male students	-.083	(.054)	-.143	(.037)***	-.080	(.029)***
Percent full-time students	.406	(.047)***	.242	(.030)***	.286	(.041)***
Log enrollment	.030	(.008)***	.039	(.006)***	.031	(.007)***
Percent white students	-.002	(.031)	.057	(.033)*	.049	(.038)
Percent black students	-.137	(.068)**	-.124	(.050)**	.036	(.055)
Percent Hispanic students	.050	(.058)	.059	(.049)	-.115	(.109)
Mainly residential	.047	(.011)***	.049	(.009)***	.019	(.015)
Admit rate	-.100	(.023)***	-.076	(.017)***	-.024	(.019)
Public	-.059	(.015)***	-.096	(.012)***	-.059	(.013)***
HBCU	.085	(.056)	.109	(.036)***	.044	(.035)
Constant	-.283	(.096)***	-.182	(.074)**	-.335	(.087)***
Number of colleges	253		526		500	

Notes:

1) * represents p<.10, ** represents p<.05, and *** represents p<.01. Standard errors are in parentheses.

2) We used multiple imputation on a small number of observations.

Although the regression coefficients are similar using one or three years of data, the precision of our value added estimates is significantly improved by using multiple years of data, especially for smaller colleges.

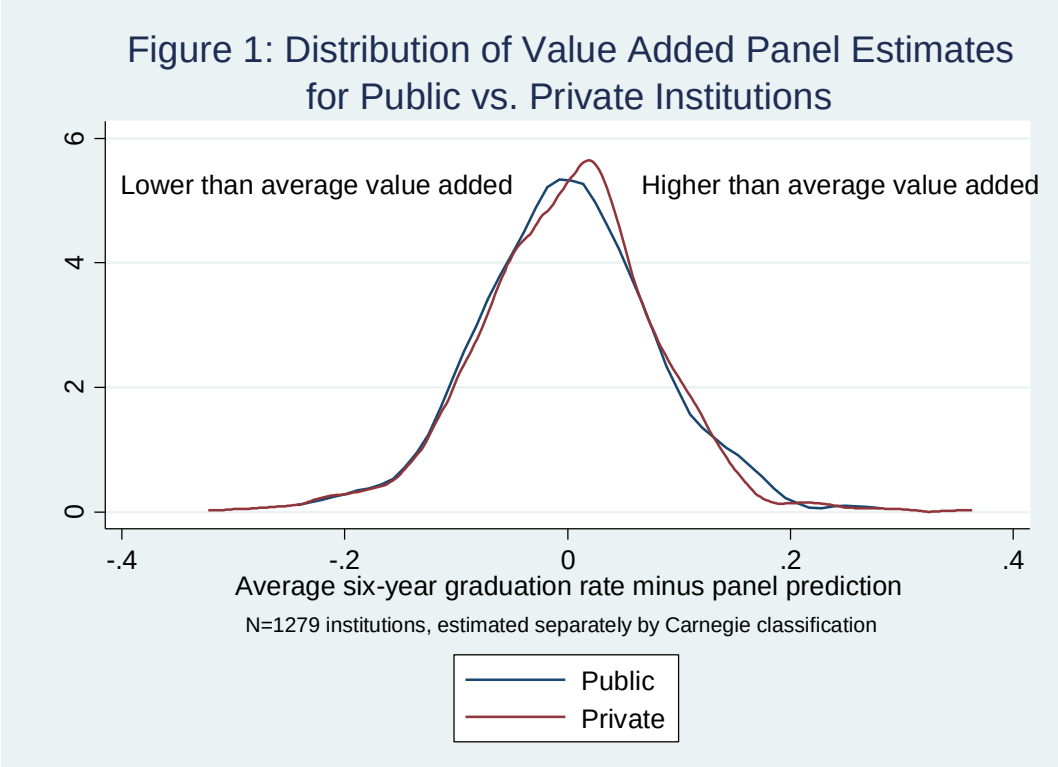
To highlight the potential significance of pooling the data, Table 4 below reports the correlation in value added measures across years, or "stability."[46] The correlation in value added just two years apart is about 0.5. One possible reason for the apparently low correlation is random error. By controlling for student characteristics, value added measures more accurately reflect the performance of institutions serving disadvantaged students, but at a cost of increasing random error.[47] The correlations between the one-year and multiyear rankings are between 0.68 and 0.84. Since the variation over time most likely reflects random error, as opposed to true differences in performance, this reinforces the importance of pooling data across years.[48]

Table 4: Stability of Value-Added Measures Over Time
(Spearman Rank-Order Correlations)

Individual Years vs. Panel Estimates (n=1279)				
Model	Panel	2008-2009	2007-2008	2006-2007
Panel	1	--	--	--
2008-2009	0.835	1	--	--
2007-2008	0.681	0.486	1	--
2006-2007	0.693	0.487	0.808	1

Note: All correlations are significantly different from zero at p<.001.

Figure 1 below plots the distribution of three-year value added measures separately for public and private institutions, using the panel regressions in Table 3 to the average six-year graduation rate for all three academic years (estimated separately by Carnegie classification).[49] The distribution is centered at zero, with a standard deviation of approximately 8 percentage points.[50] To get a sense of what this distribution means, if colleges improve from the 50th to the 84th percentile in the college value added distribution (one standard deviation), this would mean the average graduation rate would increase by an arguably large number—8 percentage points. (This, of course, assumes that the measures can be taken literally, which is probably not the case for reasons discussed later.)



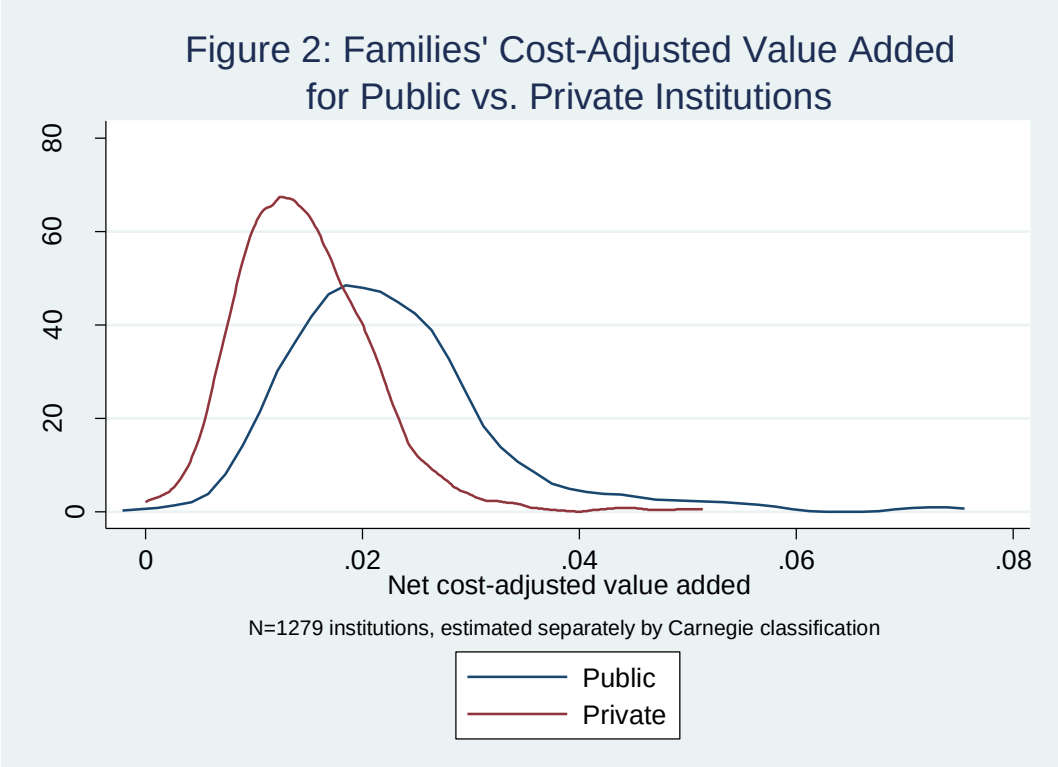
The distribution of the value added estimates is more dispersed for bachelor's-level institutions (a standard deviation of 8.5 percent) and master's-level institutions (7.9 percent) than research institutions (6.4 percent). It is important to emphasize again that those colleges with negative unadjusted value added are almost certainly not making most of their students worse off in any absolute sense; rather, it means they are below average in graduating their students (given the students' characteristics) compared with other colleges.

The distribution of estimated value added is quite similar for public and private institutions, although there are more private institutions at the extreme ends of the distribution (likely because of the smaller size of these institutions). We tested for differences in means and differences in the overall shape of the distribution and found no statistically significant differences between public and private institutions.[51]

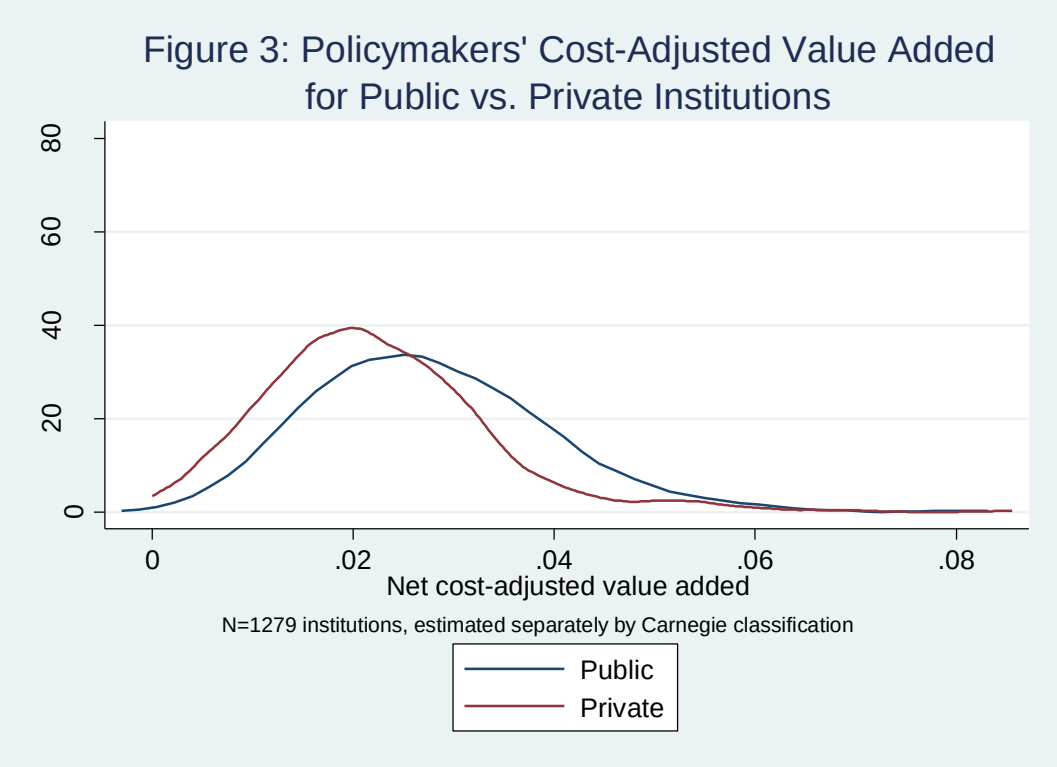
Cost-adjusted value added

In the family model, we divided the value added by the overall cost of attendance less the average amount of grant aid received by students at a given college (see Equation (3) above). We would expect public colleges to look better in cost-adjusted value added estimates because state subsidies reduce the net costs that students must pay. Additionally, in some states, students attending public colleges are eligible for additional forms of grant aid that students attending private college cannot receive. Our findings back up this

hypothesis, with public colleges having substantially higher levels of cost-adjusted value added than private colleges. Figure 2 shows the distribution of cost-adjusted value added scores by type of institutional control. Both the means and the distributions as a whole are significantly different from each other across the two sectors.



While students and their families care about the value added with respect to the net cost of attendance, policymakers are interested in the total cost of educating a student.[52] As a result, we divide the value added result by per-student educational expenses (in thousands) to get the policymaker value added estimate. Figure 3 shows the distribution of cost-adjusted policymaker value added scores by type of institutional control. As in the family model, the mean public institution has a higher value added coefficient than the mean private institution, but the difference is about half of what it was in the family model. This is what we would expect given that public resources are now added as part of costs in public colleges, which reduces cost-adjusted value added for the public colleges but not the private ones. The public-private difference is statistically significant in both models.



Comparing value added with popular ranking systems

Our next step is to see the extent to which our value added models produce sets of college rankings that are different from the popular college rankings. To do this, we compare our value added measures with the *U.S. News*, *Washington Monthly* and *Barron's* rankings using Spearman rank-order correlations separately for research universities and liberal arts colleges and present the results in Table 5. Given that *Washington Monthly* includes several different student outcomes in its measure and gives outcomes more weight than does *U.S. News*, we expected some differences between the two sets of magazine rankings. Indeed, the correlation coefficient between the two measures is 0.711 for research universities and 0.640 for liberal arts colleges. The *Washington Monthly* rankings are also somewhat less strongly correlated with the *Barron's* rankings than are the *U.S. News* rankings, which also suggests that *Washington Monthly* gives somewhat greater weight to non-input factors.

Table 5: Spearman Rank-Order Correlations Among Ranking Systems

Research Universities (n=252)

Model	*U.S. News*	*Wash. Monthly*	*Barron's*	Policymaker	Family	Overall
U.S. News	1	--	--	--	--	--
Washington Monthly	0.706	1	--	--	--	--
Barron's	0.836	0.521	1	--	--	--
Policymaker's model	-0.470	-0.217	-0.507	1	--	--
Family model	-0.176	0.080^	-0.319	0.735	1	--
Unadjusted model	0.405	0.451	0.239	0.383	0.503	1

Liberal Arts Colleges (n=220)

Model	*U.S. News*	*Wash. Monthly*	*Barron's*	Policymaker	Family	Overall
U.S. News	1	--	--	--	--	--
Washington Monthly	0.640	1	--	--	--	--
Barron's	0.833	.536	1	--	--	--
Policymaker's model	-0.570	-0.263	-0.581	1	--	--
Family model	-0.346	-0.107^	-0.415	0.685	1	--
Unadjusted model	0.448	0.438	0.274	0.233	0.322	1

Notes:
1) Rank-order correlations were only calculated for institutions with complete data in our dataset and were ranked by all three publications.
2) Institutions that were tied in the *U.S. News* rankings were assigned the median value of all the tied schools. For example, if there was a five-way tie for first place, all five institutions would be assigned a ranking of 3.
3) All correlations are statistically significant from zero at p<.01 except those marked by ^.

The correlations between those two sets of rankings and our value added rankings are much lower, however. The correlations between our (unadjusted) value added rankings and the two magazines' rankings are between 0.40 and 0.46 for both research universities and liberal arts colleges. The unadjusted model is positively correlated with the *Barron's* rankings, but less so with correlations between 0.24 and 0.27.

The correlations drop considerably in the cost-adjusted models—and even reverse direction. The rank-order correlations between the policymaker model and the three national rankings are all negative and highly significant. The correlations between the policymaker model and the *U.S. News* and *Barron's* rankings are between -0.45 and -0.60 for both types of colleges; the correlations with the *Washington Monthly* rankings are still negative, but smaller in absolute terms (around -0.25). The policymaker model and the unadjusted value added model are somewhat positively correlated (0.38 for research universities and 0.23 for liberal arts colleges). The family model has a weak negative correlation with the three national rankings and is

positively correlated with the overall and policymaker models. The negative correlations are not surprising given that increased costs necessarily lower our value added estimates, while higher costs are implicitly viewed as a positive in college rankings.

To provide a more concrete sense of how these alternative measures affect performance rankings, we list in Table 6 the top 10 colleges by Carnegie classification and model type. In the initial value added model (without adjusting for costs), there is a mix of public and private colleges in each category. Although the top-10 lists contain some colleges that are highly regarded (such as UC-San Diego, Berea College and The College of New Jersey), the lists are dominated by lesser-known institutions (such as South Carolina State University, Elizabeth City State University and Gwynedd-Mercy College).

Table 6: Top Ten Institutions by Carnegie Classification and Model

Carnegie Research (n=253)

Rank	Value-Added Model	Policymaker's Model	Family Model
1	Immaculata Univ.	Immaculata Univ.	Univ. of Florida
2	Univ. of Cal.--Davis	Univ. of Central Florida	Univ. of Georgia
3	Univ. of La Verne	Rutgers Univ.--Newark	North Carolina A&T
4	South Carolina State Univ.	Texas Woman's Univ.	South Carolina State Univ.
5	SUNY Coll. Env Sci & Forestry	South Carolina State Univ.	Univ. of Wyoming
6	Rutgers Univ.--Newark	Spalding Univ.	North Carolina State Univ.
7	Saint Louis University	Univ. of La Verne	Georgia Tech
8	Univ. of Cal.--San Diego	Mississippi State Univ.	Tennessee State Univ.
9	Univ. of Florida	Old Dominion Univ.	Univ. of Washington
10	Univ. of Washington	Jackson State Univ.	Florida State Univ.

Carnegie Master's (n=526)

Rank	Value-Added Model	Policymaker's Model	Family Model
1	The Citadel	Benedictine Univ.	CUNY John Jay Coll.
2	Coll. of Notre Dame (MD)	Towson Univ.	CUNY Brooklyn Coll.
3	Penn State--Harrisburg	Troy Univ.	The Citadel
4	Rosemont Coll.	Cardinal Stritch Univ.	NC Central Univ.
5	Gwynedd-Mercy Coll.	College of St. Joseph	Delta State Univ.
6	Coll. of Saint Elizabeth	William Carey Univ.	CUNY Lehman Coll.
7	Univ. of IL--Springfield	Gwynedd-Mercy Coll.	Albany State Univ.
8	Regis College	CUNY Baruch Coll.	CUNY Baruch Coll.
9	Rutgers Univ.--Camden	IN Wesleyan Univ.	Univ. of Texas-Pan American
10	The Coll. of New Jersey	The Citadel	CUNY Queens College

Carnegie Bachelor's (n=500)

Rank	Value-Added Model	Policymaker's Model	Family Model
1	Saint Joseph's Coll. (NY)	Ohio Christian Univ.	Elizabeth City State Univ.
2	Coll. of Our Lady of the Elms	Penn State--Dubois	OK Panhandle State Univ.
3	Penn State--Dubois	Central Baptist Coll.	US Military Academy
4	Ohio Christian Univ.	Coll. of Our Lady of the Elms	US Naval Academy
5	Berea Coll.	Dakota State Univ.	CUNY York Coll.
6	Moravian College	Penn State--Schuylkill	Virginia Military Institute
7	ME Maritime Academy	Albertus Magnus Coll.	US Air Force Academy
8	Elizabeth City State Univ.	Saint Joseph's Coll. (NY)	US Coast Guard Academy
9	CA Maritime Academy	Penn State--Mont Alto	Rust Coll.
10	Paul Quinn College	West Liberty Univ.	Paul Quinn Coll.

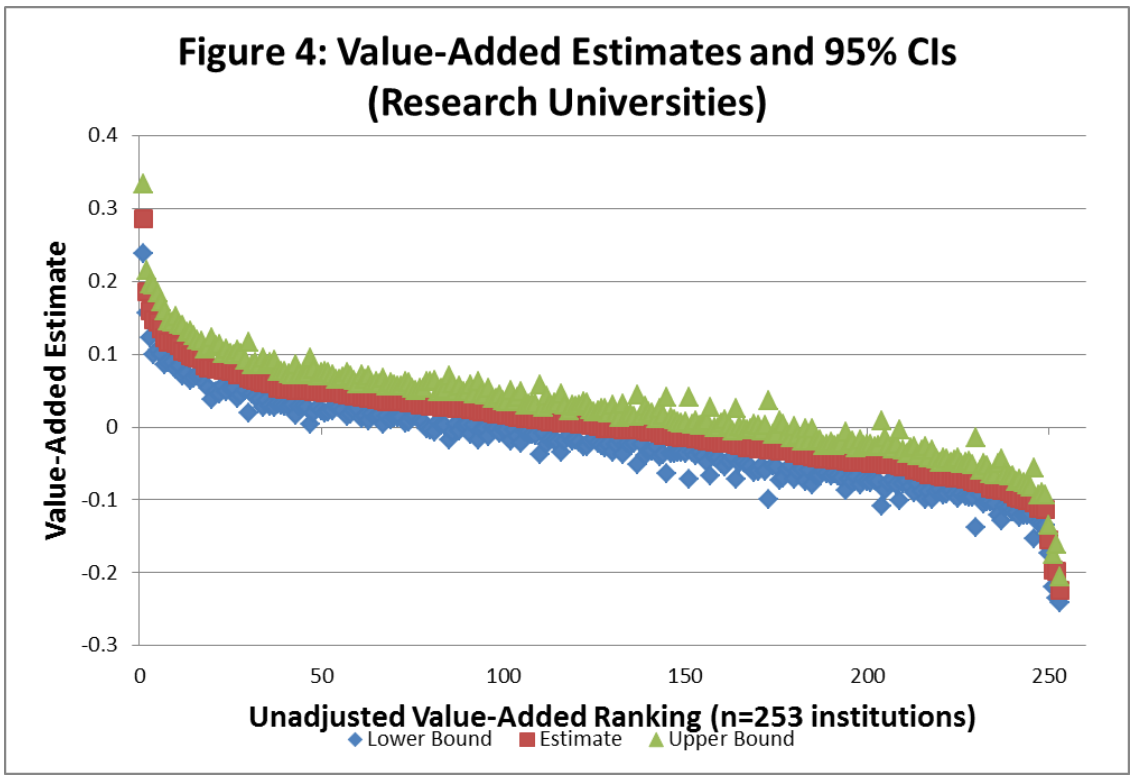
The top-10 lists change substantially for the family and policymaker models compared with the overall model. There are more public institutions in the top-10 lists for the family model than the policymaker

model, which comes as little surprise since students and parents pay a smaller portion of total costs in public colleges versus private ones. While many institutions make two of the three top-10 lists, only two appear on all three lists: South Carolina State University (research) and The Citadel (master's). Although these institutions could be considered exemplary, we know little about the reasons behind the high value added coefficients. They should be studied in more detail to see whether their high value added coefficients are a result of institutional practices or random error.

It is worth emphasizing the differences between our top-10 lists and the results from the popular college rankings. For example, South Carolina State University (ranked third in the unadjusted model, fifth in the policymaker's model and fourth in the family model for research universities) is ranked sixth in the *Washington Monthly* national university rankings (and first in outperforming its predicted graduation rate), while it languishes in the third tier (ranked 134-190) in the *U.S. News* national university rankings and is classified as "less competitive" by *Barron's*, the second-lowest ranking.

In his critique of the *U.S. News* rankings, Stephen Porter noted that value added estimates of graduation rates should be examined to see whether the estimate is significantly different from zero.[53] We used the panel regression estimates separately by Carnegie classification (from Table 3) and estimated 95 percent confidence intervals for the unadjusted model. With the additional years of data, we are able to estimate more precise confidence intervals and hence more institutions have statistically significant value added. For research institutions, 33.6 percent have value added estimates that are significantly greater than average (zero), and 32.8 percent have value added estimates that are significantly less than average; the corresponding numbers for master's and bachelor's institutions are 36.7 percent vs. 36.9 percent and 39.8 percent vs. 36.6 percent, respectively. The remainder of institutions have value added that is not significantly different from average.

We also tested the sensitivity of the rankings to random error. Figure 4 presents 95 percent confidence intervals for the unadjusted value added model for all research institutions.



This figure suggests that institutions ranked within approximately 30 to 100 places of each other have unadjusted value added estimates that are not statistically significant from each other. An example of how a top-10 institution could be affected by measurement error is the University of Florida, which ranks ninth on the unadjusted model for research institutions. Its confidence interval encompasses the confidence intervals of institutions ranked from third to 35th. This suggests that an institution's precise ranking is inexact, but its general position on the distribution is only somewhat affected by the presence of measurement error.

Discussion and limitations

Value added measures are fairly insensitive to the different approaches we have considered here. However, there are four main limitations of these data: (1) We have only one student outcome at one time point; (2) we have data only at the college level; (3) we have somewhat limited information about student background; and (4) the measures are potentially problematic for institutions with extremely high graduation rates.

First, unlike in K-12 value added research, which involves annual measures of student test scores for each student, we do not—and cannot—have annual measures for individual students. Students can go only from "not graduated" to "graduated." It is not clear whether this is an issue. Certainly having prior information on the same outcome for the same students greatly increases our ability to predict educational outcomes, but it is still possible that the relative position of colleges is insensitive to this.

In higher education, an instrument such as the Collegiate Learning Assessment (CLA), which seeks to measure students' critical thinking and reasoning knowledge, could be used to institute a form of value added very similar to K-12.[54] However, it seems likely that the CLA is less well aligned with the intended content of college instruction than is the case in K-12 schools with their explicit content standards and increasing alignment with state tests.

We cannot emphasize enough the multiple objectives of education. Focusing on graduation at the exclusion of educational quality not only provides a potentially misleading picture but also opens up the possibility of perverse incentives. If colleges are given incentives to produce more degrees, they can easily do so by simply making graduation easier through lowering standards or steering students toward certain "easy" majors.[55] It would also be helpful to have data on employment, earnings, health and other longer-term outcomes.[56]

Second, while it would seem that having student-level data is much better, the advantages are somewhat unclear. On the positive side, having data on individual students would allow us to identify effects on subgroups (e.g., adding part-time and non-first-time students whom we are forced to exclude by using IPEDS) and allow us to define outcomes in different ways (e.g., treating lateral transfers differently from dropouts). Student-level data also make it possible to provide a more realistic assessment of the statistical significance of the estimates. Aggregated data tend to yield a false sense of precision by ignoring the fact that certain types of students end up in certain colleges.

However, some research suggests that aggregated data lead to upwardly biased estimates of the role of school inputs, although this may or may not apply to higher education or the effects of entire institutions.[57] Aggregated data incorporate both the effects of certain background characteristics for each individual student and the ways in which groupings of students affect overall institutional performance. For example, while student ability (as measured by standardized test scores) affects each individual student's likelihood of graduating, high concentrations of lower-ability students may further reduce a student's likelihood of graduation.[58]

A third and possibly more important issue is the richness of the variables intended to capture students' backgrounds, but this issue is separate from data aggregation. The IPEDS data, with information about ACT/SAT scores and financial aid receipt, may be richer than some student-level data sets created by state governments.[59] Ideally, we would have student-level data with rich covariates and we could test the sensitivity of results to data aggregation and the availability of student background factors.

Fourth, our measures are potentially problematic for institutions with extremely high graduation rates. None

of the top 10 colleges in our value added measures are also among the top 10 in the *U.S. News* rankings. In fact, among the top 10 research institutions in the *U.S. News* rankings, the highest-ranked in our ratings is Harvard, which ties for first in the *U.S. News* rankings but ranks 12th of 250 in the unadjusted value added model, 230th in the policymaker model and 138th in the family model. This is not a surprising result because by privileging resources and prestige over institutional effectiveness, the rankings create a systematic disadvantage for colleges serving lower-income and lesser-prepared students.[60]

Because we have reason to suspect that the model might not work well for schools with very high actual graduation rates, we re-estimated the unadjusted value added models, excluding the top 30 institutions in the *U.S. News* research university and liberal arts college rankings. This exclusion did not meaningfully affect the rankings of the institutions ranked lower by *U.S. News*. The correlation in rankings between these two approaches for the colleges that are in both analyses is 0.99 for research universities and 0.94 for liberal arts colleges. We also used a nonlinear model and obtained even higher correlations.[61]

Possible alternative cost-adjusted value-added measures

One of our arguments is that costs and contributions to outcomes should both be part of any performance metric. But this is not necessarily the only way to do it. To see why we simply divided unadjusted value added by costs, it is important to consider the alternatives.

In theory, we could quantify the dollar value of the benefits of a college degree and of dropping out (or transferring from a given institution). Unfortunately, this is impossible here because we would need valid estimates of the economic and other returns to earning a degree at each college. More states are building the capacity for such an analysis. But this cannot be done on anything close to a national level, nor for the wide variety of noneconomic benefits that education affects.[62]

Even if we could estimate those economic returns, a simple benefit-cost ratio such as ours above is still a useful metric. Suppose that families act the way economists assume: They make educational decisions to maximize their happiness or utility. The optimal solution is where the happiness generated by an additional year of education (divided by the price of education) equals the additional happiness generated by other investments and/or consuming other goods (again, divided by prices). Moreover, economists generally assume that additional income increases happiness at a diminishing rate, so that going from $10,000 to $20,000 in annual income increases well-being more than going from $100,000 to $110,000. For this reason, we might think that low-income and high-income families are likely to view the costs and benefits of college differently. For example, it is likely that the children of high-income families will have greater wealth regardless of their education level (e.g., because of gifts and inheritance from parents), and this may lead high-income families to worry less about the income generated by degrees. This also means that high-

income families will also be less worried about the costs, and so these differences will partially cancel out when we calculate the ratio. Thus, while we do not question that low-income and high-income families look at college decisions differently, our simple ratio is still broadly useful.

Another alternative would be to just keep the unadjusted value added and cost information separate. This might encourage families and policymakers to make their own judgments about how important each part is. If high-income families are willing to pay whatever is necessary to obtain the best education for their children, they might wish to ignore costs altogether. But very few families are in this situation. We argue that the vast majority of families want to weigh college performance in relation to costs, but the information they currently receive makes it difficult to do so. No single metric can account for the diverse preferences, incomes and other constraints that families face, but the ratio of unadjusted value added to costs has definite advantages over the alternatives.

Our cost-adjusted value added measure is also intuitive: Other things equal, producing more actual graduates improves the measure, while increasing cost reduces it. Increasing the predicted graduation rate, other things equal, reduces the value added measure because this higher prediction implies that graduation is being driven by student characteristics rather than institutional effectiveness. Consider the following example. Institution A has an unadjusted value added of 0.10 and spends $10,000 per student, while Institution B has an unadjusted value added of 0.15 and spends $30,000 per student. Institution A therefore has a cost-adjusted value added of 0.01 (recall that we measure costs in thousands), while Institution B has a cost-adjusted value added of 0.005. In order to be as cost-effective as A, Institution B would have to either increase its graduation rate by 15 percentage points or cut its costs in half.

But in other respects, the measure is hard to interpret. When we say that someone is 6 feet tall, we know what a foot means. But what does "cost-adjusted value added" mean? This is not an easy question to answer. In our view, it is best to view cost-adjusted value added as an index, much like the Dow Jones Industrial Average. What does it mean when the Dow drops by 100 points? There is no simple average, yet it is still viewed as a useful measure of the health of the stock market as a whole. Popular college rankings such as *U.S. News* are also based on indices. As we described in the second section, each component (e.g., peer assessment and average ACT/SAT scores) is given a certain amount of weight and scaled in a way that boils down all the information into a single number that is used to rank colleges. We are doing the same here.

Recommendations and conclusions

Different stakeholders necessarily mean different things by "college performance," but our results suggest that the information currently available is inadequate for everyone. Students, parents and policymakers have been taught to believe that the best colleges are the most prestigious and expensive ones, attended by the

most academically prepared students. Rather than encouraging efficiency, this motivates colleges to attract better students and continually raise costs and tuition—but not to use resources more wisely. This is one of the major causes behind both the college cost spiral and the stagnant graduation rates over the past two decades.[63] The information problem yields little incentive to either reduce costs or improve graduation rates and therefore impedes progress toward the increasingly high goals we have for higher education.

It is doubtful that existing college ranking producers will make changes to address the major flaws in their systems. One option would be to produce an entirely new ranking system, along the lines of what we propose here (and improved upon with additional research). Such a system would have to heavily emphasize that the measures are intended only as information and that no single metric could possibly account for the vast array of factors that go into students' college decisions. We would oppose tying the measures to high-stakes decisions, given that we know so little about them at this stage and that only one outcome (graduation) is measured.

It is only a matter of time until it will be possible to estimate value added to longer-term outcomes such as employment for large numbers of colleges nationwide. These additional outcomes could easily be incorporated into the framework we have presented. The unadjusted value added would become a weighted average of the value added estimated using different outcome measures, and these separate measures could be combined and then adjusted for costs. We believe the main problem with the value added measures we have presented here is that they focus only on graduation—and lack even an indirect measure of quality—so incorporating additional outcomes would be a major advance.

It will also be increasingly feasible to estimate value added for different subgroups. We argue that different families have different needs—and they certainly have different rates of success in college. The more we can target information to specific families, the more useful it is likely to be. Compared with college rankings, the basic value added approach is likely to be more useful to a much broader range of constituencies. Low- and middle-income families often do not have the luxury of worrying about prestige and instead simply want to make sure they will be able to get a degree and a quality education.

Value added measures might also be a basis for learning about successful programs and practices. What distinguishes low- and high-value added colleges? The answer to this question might be more important to improving student success than a formal accountability system. Moving from the 50th to the 84th percentile on our value added measure is associated with an increase of 8 percentage points in the graduation rate. If those differences reflect different strategies, and those strategies could be identified and scaled up, the graduation rate might be improved noticeably.

Value added could also help resolve the current policy debate surrounding for-profit colleges. Taking advantage of recent changes in federal financial aid laws that make it easier for students to use federal aid at these institutions, the for-profit sector has grown dramatically. Critics point to low loan repayment rates and allege misleading advertising and shallow instructional programs, while the institutions themselves argue that they serve some of the most disadvantaged students with more flexible offerings and greater efficiency. How can policymakers ensure that for-profit (and not-for-profit) institutions are making wise use of public resources? Value added is one possibility. With more disadvantaged students, even high-performing colleges will have lower graduation rates—whether they are for-profit or nonprofit. Policymakers need to find a way to reward these colleges for their successes while taking into account the additional hurdles their students face. We cannot estimate the value added of for-profits with the data we have, but such data collection might be part of the public policy plan for holding all institutions accountable.

We can never measure college success perfectly, and failing to recognize performance measurement problems can have serious consequences. But there are better ways to use information and improve performance. Although we fully acknowledge the limitations of our rankings, we argue that using value added to estimate the cost-effectiveness of colleges is a step forward.

Table A1: Regression Results by Year (All Institutions)

	2008-09		2007-08		2006-07	
Variable	Coeff	(SE)	Coeff	(SE)	Coeff	(SE)
Percent Pell	-.240	(.035)***	-.395	(.036)***	-.442	(.033)***
Average ACT composite	.028	(.001)***	.026	(.002)***	.025	(.001)***
Percent receiving financial aid	-.111	(.028)***	-.049	(.028)*	-.053	(.024)**
Percent receiving student loan	.136	(.021)***	.096	(.021)***	.106	(.019)***
Percent male students	-.087	(.024)***	-.131	(.025)***	-.133	(.023)***
Percent full-time students	.231	(.026)***	.286	(.028)***	.298	(.025)***
Log enrollment	.025	(.005)***	.025	(.005)***	.021	(.005)***
Percent white students	.008	(.027)	-.011	(.027)	-.010	(.025)
Percent black students	-.116	(.039)***	-.048	(.043)	.023	(.040)
Percent Hispanic students	.081	(.045)*	.076	(.047)	.073	(.044)*
Carnegie research	.007	(.010)	.000	(.011)	.005	(.010)
Carnegie master's	.005	(.007)	-.005	(.008)	-.002	(.007)
Mainly residential	.035	(.008)***	.029	(.008)***	.024	(.008)***
Admit rate	-.091	(.018)***	-.076	(.019)***	-.055	(.017)***
Public	-.054	(.009)***	-.047	(.009)***	-.042	(.009)***
HBCU	.083	(.026)***	.093	(.028)***	.069	(.026)***
Constant	-.299	(.069)***	-.283	(.067)***	-.245	(.061)***
Number of colleges	1279		1279		1279	
Adjusted R-squared	.739		.719		.741	

Notes:

1) * represents p<.10, ** represents p<.05, and *** represents p<.01. Standard errors are in parentheses.

2) We used multiple imputation on a small number of observations.

Table A2: Regression Results by Carnegie Classification

	Research		Master's		Bachelor's	
Variable	Coeff	(SE)	Coeff	(SE)	Coeff	(SE)
Percent Pell	-.170	(.100)*	-.162	(.058)***	-.359	(.056)***
Average ACT composite	.028	(.003)***	.025	(.003)***	.030	(.002)***
Percent receiving financial aid	-.119	(.048)**	-.192	(.054)***	-.018	(.053)
Percent receiving student loan	.088	(.042)**	.171	(.033)***	.101	(.036)***
Percent male students	-.234	(.064)***	-.106	(.045)**	-.048	(.035)
Percent full-time students	.373	(.055)***	.214	(.038)***	.206	(.052)***
Log enrollment	.015	(.009)*	.024	(.008)***	.021	(.009)**
Percent white students	-.039	(.044)	-.041	(.046)	.083	(.050)
Percent black students	-.210	(.089)**	-.177	(.064)***	.021	(.069)
Percent Hispanic students	.002	(.074)	.039	(.066)	-.053	(.132)
Mainly residential	.036	(.013)***	.054	(.011)***	-.002	(.018)
Admit rate	-.070	(.037)*	-.114	(.028)***	-.088	(.031)***
Public	-.021	(.018)	-.063	(.015)***	-.053	(.016)***
HBCU	.102	(.063)	.066	(.043)	.075	(.041)*
Constant	-.258	(.137)*	-.145	(.111)	-.389	(.121)***
Number of colleges	250		516		475	
Adjusted R-squared	.843		.613		.740	
Average graduation rate (%)	64.6		51.1		54.5	

Notes:

1) * represents p<.10, ** represents p<.05, and *** represents p<.01. Standard errors are in parentheses.

2) We used multiple imputation on a small number of observations.

3) Results are reported for the 2008-09 academic year.

[1] The amount of financial aid that follows a student to a given institution depends on institutional factors such as the cost of attendance as well as state policies.

[2] Obama, Barack. "Remarks by the President on the American Graduation Initiative." *The White House*. July 12, 2009. http://www.whitehouse.gov (accessed July 26, 2009).

[3] Lumina Foundation for Education. A Stronger Nation Through Higher Education. Indianapolis, Indiana: Lumina Foundation for Education, 2009.

[4] The graduation rate reported here and throughout this paper is for first-time, full-time students who begin at a four-year university and fails to reflect the large and growing number of nontraditional and community college students. John Bound, Michael F. Lovenheim, and Sarah Turner. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources." American Economic Journal: Applied Economics 2, no. 3 (July 2010): 129-157.

[5] Douglas N. Harris and Sara Goldrick-Rab. "The (Un)Productivity of American Colleges: From "Cost Disease" to Cost-Effectiveness." University of Wisconsin-Madison, Madison, 2010.
[6] William J. Baumol and Sue Anne Batey Blackman. "How to Think About Rising College Costs." Planning for Higher Education 23, no. 4 (Summer 1995): 1-7. Howard R. Bowen. The Costs of Higher Education: How Much Do Colleges and Universities Spend per Student and How Much Should They Spend? San Francisco, California: Jossey-Bass, 1980. David W. Breneman, "An Essay on College Costs." In Study of College Costs and Prices, 1988-89 to 1997-98, by Alisa F. Cunningham, et al, 13-20. Washington, DC: U.S. Department of Education, National Center for Education Statistics, 2001. Ronald G. Ehrenberg. Tuition Rising: Why College Costs so Much. Cambridge, Massachusetts: Harvard University Press, 2000. Malcolm Getz and John J. Siegfried. "Cost and Productivity in American Colleges and Universities." In Economic Challenges in Higher Education, edited by Charles Clotfelter et al, 261-392. Chicago, Illinois: University of Chicago Press, 1991. Harris and Goldrick-Rab. "The (Un)Productivity of American Colleges." Richard Vedder. Going Broke by Degree: Why College Costs Too Much. Washington, DC: AEI Press, 2004. Burton A. Weisbrod, Jeffrey, P. Ballou, and Evelyn D. Asch. Mission and Money: Understanding the University. New York: Cambridge University Press, 2008.
[7] Bobbi L. Wolfe and Robert H. Haveman. "Social and Nonmarket Benefits from Education in an Advanced Economy." Boston Federal Reserve Conference Series, June 2002: 97-142.
[8] Peter Ewell. "Developing Statewide Performance Indicators for Higher Education: Policy Themes and Variations." In Charting Higher Education Accountability: A Sourcebook on State-Level Performance Indicators, edited by Sandra Ruppert, 83-93. Denver, Colorado: Education Commission of the States, 1994.
[9] Michael McLendon, James Hearn, and Russ Deaton. "Called to Account: Analyzing the Origins and Spread of State Performance-Accountability." Educational Evaluation and Policy Analysis 28, no. 1 (Spring 2006): 1-24.
[10] U.S. Department of Education. A Test of Leadership: Charting the Future of U.S. Higher Education. Washington, 2006.
[11] Andrew P. Kelly and Mark Schneider. Filling in the Blanks: How Information Can Affect Choice in Higher Education. Washington, DC: American Enterprise Institute, January 2011.
[12] Thomas Bailey et al. "Is Student Right-to-Know All You Should Know? An Analysis of Community College Graduation Rates." Research in Higher Education 47, no. 5 (August 2006): 491-519.
[13] Kevin Carey and Andrew P. Kelly. The Truth Behind Higher Education Disclosure Laws. Washington, DC: American Enterprise Institute, November 2011.
[14] Patricia McDonough et al. "College Rankings: Democratized College Knowledge for Whom?" Research in Higher Education 39, no. 5 (1998): 513-537.
[15] James Monks and Ronald Ehrenberg. The Impact of U.S. News and World Report College Rankings on Admissions Outcomes and Pricing Policies at Selective Private Institutions. Working Paper 7227, Cambridge: National Bureau of Economic Research, 1999. Marc Meredith. "Why Do Universities Compete in the Ratings Game? An Empirical Analysis of the Effects of the U.S. News and World Report College Rankings." Research in Higher Education 45, no. 5 (August 2004): 443-461. Nicholas Bowman and Michael Bastedo. "Getting on the Front Page: Organizational Reputations, Status Signals, and the Impact of U.S. News and World Report on Student Decisions." Research in Higher Education 50 (2009): 415-436.
[16] At national universities and liberal arts colleges, the combined graduation and retention rates are worth 20% and graduation rate performance is worth 5%. At master's and baccalaureate universities, the combined graduation and retention rates are worth 25% and no weight is given to graduation rate performance. Robert Morse and Sam Flanigan. "How We Calculate the College Rankings." U.S. News and World Report, August 19, 2009.
[17] Whether the peer assessment measure reflects inputs or outputs is more ambiguous. We discuss the "graduation rate performance" measure later in the paper.
[18] In the peer assessment and alumni giving measures, it is possible that inputs indirectly reflect an institution's operating efficiency.
[19] Catherine Watt. "Strategic Planning: Project Management and Lessons in Moving up the U.S. News Rankings." 2009 Association for Institutional Research Conference. Atlanta, 2009.
[20] Barron's Educational Series, Inc. Barron's Profiles of American Colleges: Descriptions of the Colleges. Hauppauge, New York: Barron's Educational Series, Inc., 2009.
[21] Scott E. Carrell, Richard L. Fullerton, and James E. West. "Does Your Cohort Matter? Peer Effects in College Achievement." Journal of Labor Economics 27, no. 3 (July 2009): 439-464. Ralph Stinebrickner and Todd R. Stinebrickner.

"What Can Be Learned About Peer Effects Using College Roommates? Evidence From New Survey Data and Students From Disadvantaged Backgrounds." Journal of Public Economics 90 (2006): 1435-1454.
[22] Jennie E. Brand and Charles N. Halaby. "Regression and Matching Estimates of the Effects of Elite College Attendance on Educational and Career Achievement." Social Science Research 35 (2006): 749-770.
[23] Jacques Steinberg. "Is Going to an Elite College Worth the Cost?" The New York Times, December 17, 2010: WK1. Stacey Dale and Alan B. Krueger. Estimating the Return to College Selectivity over the Career Using Administrative Earning Data. Princeton, NJ: Industrial Research Section, February 2011. Also note that Dale and Krueger found that there are significant positive returns to college selectivity for certain disadvantaged subgroups.
[24] The "arms race" for the best and brightest students can be present even with a focus on increasing the representation of historically underrepresented students. For example, institutions compete for the most talented low-income and minority students so they can satisfy both the equity and excellence agendas.
[25] Mause , in his "excessive signaling hypothesis," contended that this competition is detrimental to universities, as they expend a large amount of resources in a few categories such as targeted scholarships and campus recreation. Karsten Mause. "Too Much Competition in Higher Education? Some Conceptual Remarks on the Excessive-Signaling Hypothesis." American Journal of Economics and Sociology 68, no. 5 (November 2009): 1107-1133.
[26] Harris and Goldrick-Rab. "The (Un)Productivity of American Colleges."
[27] Colleges do control their student bodies through the admissions process, but this aspect of control is unrelated to what is typically thought of as performance.
[28] Douglas N. Harris. Value Added Measures in Education: What Every Educator Needs to Know. Cambridge, Massachusetts: Harvard Education Press, 2011. Douglas N. Harris. "Would Accountability Based on Teacher Value Added be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives." Education Finance and Policy 4, no. 4 (2009): 319-350.
[29]It is important to note that the predicted graduation rates are sensitive to the inclusion or exclusion of a few colleges in the rankings. As such, the sample of colleges must be defined carefully when generating predicted graduation rates. Stephen Porter. "The Robustness of the "Graduation Rate Performance" Indicator Used in the U.S. News and World Report College Rankings." 2nd AIR-CASE Conference. Washington, 1999.
[30] In the rankings released in the fall of 2010, the weight given to graduation rate performance increased from 5 to 7.5 percent.
[31] Averaging the25th and 75th percentile may or may not yield the average score, depending on the distribution of scores in each institution; however, this is a standard way to summarize ACT/SAT scores and there is no obvious alternative short of reporting the two separately. The weight given to this value added measure in the *Washington Monthly* rankings is not published on the magazine's website: http://www.washingtonmonthly.com/college_guide/feature/a_note_on_methodology_1.php.
[32] We use the average ACT score, while the *Washington Monthly* uses the average SAT score. Since we convert SAT scores into ACT scores using the official concordance, there should not be a meaningful difference in the results due to test scores. College Board. "SAT Score Use Practices." September 2009. http://www.professionals.collegeboard.com/profdownload/sat-score-use-practices-list.pdf.
[33] More information can be found at http://www.collegeresults.org/aboutthedata.aspx.
[34] Data for the percentage of students receiving Pell Grants, financial aid, and student loans is provided courtesy of The Institute for College Access & Success. College InSight, http://college-insight.org. Student debt and undergraduate financial aid data are licensed from Peterson's Undergraduate Financial Aid and Undergraduate Databases, © 2010 Peterson's, a Nelnet company, all rights reserved.
[35] A small percentage of institutions are missing data on certain other variables; we use multiple imputation to estimate these observations. Also, note that in the past, IPEDS has only reported the percentage of first-time, full-time undergraduate students receiving Pell Grants. Although more institutions report these data to IPEDS than the overall percentage of undergraduates receiving Pell Grants (taken from College InSight for 2006-07 and 2007-08), we feel that using the more inclusive definition of undergraduates is important.
[36] We specifically use Ordinary Least Squares (OLS) as our baseline approach, although we also try other approaches as noted later.

[37] We do include the percentage of students receiving any financial aid and the percentage of students receiving loans in our regression models, even though these factors may be partially within an institution's control. Estimating our models without these two measures does not substantively change our results.
[38] A case can be made for excluding all institutional characteristics based on the above reasoning that parents and policymakers have reasons to make comparisons across institutional types. On the other hand, unobservable differences across students may be correlated with these institutional characteristics. Therefore we include them in the baseline model, but in future work will consider alternative models.
[39] We also estimated models that included research expenses in the educational expenses category. This did not substantively change our findings.
[40] We specifically estimate the panel regression using random effects and maximum likelihood estimation.
[41] Douglas N. Harris. "Toward Policy-Relevant Benchmarks for Interpreting Effect Sizes: Combining Effects with Costs." Educational Evaluation and Policy Analysis 31, no. 1 (2009): 3-29. Henry M. Levin and Patrick J. McEwan. Cost-Effectiveness Analysis (2nd Ed.). Thousand Oaks, CA: Sage Publications. 2001.
[42] We are also well aware of the role of politics in these decisions, but the point of this paper is to develop a measure that provides a more objective and informative indication of performance.
[43] Note that IPEDS does not have information at the department level, so we can only estimate the effectiveness-cost ratio for an entire institution.
[44] We would like to be able to estimate the net cost of attendance for individual students, but this would require individual-level data about a student's financial aid package and family income. This would also allow us to estimate the marginal utility of income, which could help explain how families choose between more-expensive and less-expensive institutions.
[45] Using data from the 2006-07 or 2007-08 academic years does not significantly change the regression results. We present the results for all three years in Table A1 of the Appendix.
[46] See, for example, Koedel and Betts, who study the stability of teacher value added measures. Cory Koedel and Julian R. Betts. Re-Examining the Role of Teacher Quality in the Educational Production Function. Working Paper 07-08, Columbia, Missouri: University of Missouri-Columbia, 2007.
[47] Harris. "Value Added Measures in Education."
[48] Although standard errors are not reported in popular rankings, Dichev shows that between 70 and 80 percent of the annual *change* in rankings is a result of random and systemic error that have nothing to do with changes in college quality. As a result, Clarke contended that placing colleges in similar groups would be more accurate than a rank order when college rankings are not statistically significant from each other, though even this would still mean that colleges near the cutoff scores would be frequently placed in the wrong groups. Ilia Dichev. "News or Noise? Estimating the Noise in the U.S. News University Rankings." Research in Higher Education 42, no. 3 (2001): 237-266. Marguerite Clarke. "News or Noise? An Analysis of U.S. News and World Report's Ranking Scores." Educational Measurement: Issues and Practice 21, no. 4 (Winter 2002): 39-48.
[49] The kernel density plots by Carnegie classification are available from the authors upon request.
[50] Although the value added distribution is centered at zero, this does not mean that students are harmed by attending an institution with negative value added (an actual graduation rate lower than the predicted rate). The majority of students attending these institutions are still better off than they would have been had they not attended college at all.
[51] We used a Kolmogorov-Smirnov test to test the equality of distributions.
[52] Policymakers should not care about the cost of room and board because this money is kept separate from academic expenses. Public funds (in terms of institutional appropriations) do not fund room and board.
[53] Porter. "The Robustness of Graduation Rate Performance."
[54] Stephen Klein et al. "The Collegiate Learning Assessment: Facts and Fantasies." Evaluation Review 31, no. 5 (October 2007): 415-439.
[55] This is a limitation of all ranking systems (including *U.S. News* and *Washington Monthly*) that use graduation rates as a component.
[56] Anthony P. Carnevale, Michelle Melton, and Laura Meyer. "Missing the Point." Inside Higher Ed. August 26, 2010. http://www.insidehighered.com/views/2010/08/26/carnevale.

[57] Eric A. Hanushek, Steven G. Rivkin, and Lori L. Taylor. "Aggregation and the Estimated Effects of School Resources." The Review of Economics and Statistics 78, no. 4 (November 1996): 611-627.
[58] Scott E. Carrell, Richard L. Fullerton, and James E. West. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." Journal of Labor Economics 27, no. 3 (July 2009): 439-464. David J. Zimmerman. "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment." The Review of Economics and Statistics 85, no. 1 (February 2003): 9-23.
[59] We are most familiar with the Florida K-20 Education Data Warehouse, which is considered one of the best unit record data systems in the country. Although it has a considerable amount of information on a student's progression through K-12 education, it does not have financial aid data. This makes estimating the cost-adjusted model for families difficult.
[60] Harris makes the same observation about K-12 schools. Harris. "Value Added Measures in Education."
[61] Specifically, we used a generalized linear model with a logit link, which limits the predicted graduation rate between zero and one.
[62] Wolfe and Haveman. "Social and Nonmarket Benefits from Education."
[63] Harris and Goldrick-Rab. "The (Un)Productivity of American Colleges."