INPUT-ADJUSTED GRADUATION RATES AND COLLEGE ACCOUNTABILITY: WHAT IS KNOWN FROM TWENTY YEARS OF RESEARCH?

Thomas Bailey and Di Xu

September 2012

HCM STRATEGISTS

CCRC Community College Research Center
TEACHERS COLLEGE, COLUMBIA UNIVERSITY

ABSTRACT

Evaluating college performance is a complex task. The inherent multidimensionality of higher education makes it difficult to observe the entire package of knowledge and skills imparted to students, and wide variations across institutions in both their educational agendas and student profiles make it difficult to compare institutions. This paper reviews research aimed at measuring and comparing the quality and effectiveness of colleges and universities. We first review possible measures of education quality and the history of college rankings. We then present arguments for adjusting raw graduation rates for institutional and student characteristics as well as for what inputs should be adjusted. We discuss the advantages and disadvantages of methods used to adjust graduation rates and shed light on the particular difficulties of measuring graduation rates in the community college setting. We conclude with a summary of findings and recommendations.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

TABLE OF CONTENTS

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

1. INTRODUCTION

Higher education has enjoyed a long period of high prestige and general public support. Throughout the latter decades of the 20th century, Americans generally believed that their colleges and universities were producing an excellent product. But over the past 10 or 15 years, that sparkling reputation has faded, and students, parents, policymakers and taxpayers have all begun to question the quality of American higher education. Increasingly harsh fiscal realities have also led to sharp reductions in state subsidies for public colleges and universities. Since the late 1990s, consistent data on graduation rates for most colleges have shown that many of them graduate fewer than half of their students, and more recently, embryonic data on student learning of general academic skills suggest that many students who do get through college may not learn very much (Arum & Roksa, 2010). This has occurred at the same time that the federal government and prominent foundations have begun to call for significant increases in the number of individuals with college degrees (Bailey, 2011).

Despite continued high economic returns to a college education, these trends have led to increased interest in measuring the quality of higher education and a growing desire for federal and state governments to hold higher education to new standards of accountability. States are the primary funders of public institutions. The large majority of state funding is based on enrollments, but many states have tried to develop systems to reward colleges for performance, rather than simply enrollments. At the beginning of the past decade, according to the American Association of State Colleges and Universities (2002), 41 states used graduation rate data for state-level accountability and performance reporting purposes, and seven states directly employed this indicator in performance funding systems. This trend has continued during the past few years, although so far these systems have not involved large percentages of college funding (Dougherty, Natow, Hare, Jones, & Vega, 2011; Shulock & Jenkins, 2011).

Although interest in the matter has been growing, evaluating college performance is a complex task. The inherent multidimensionality of higher education has made it difficult to observe the entire package of knowledge and skills imparted to students; in the meantime, the wide variations across institutions in both their educational agendas and student profiles make it difficult to compare institutions.

This paper reviews research aimed at measuring and comparing the quality and effectiveness of colleges and universities. In Section 2 we review possible measures of education quality and the history of college rankings. Section 3 presents major arguments for adjusting raw graduation rates for institutional and student characteristics as well as for what inputs should be adjusted. Section 4 reviews major methods used to adjust graduation rates and discusses the advantages and

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

disadvantages of each one. Section 5 discusses the particular difficulties of measuring graduation rates in the community college setting. Section 6 presents a summary of this literature review and offers recommendations.

2. THE HISTORY OF EVALUATING COLLEGE PERFORMANCE

2.1 Evaluating College Performance: Which Proxy Should We Use?

The assessment of education quality actually started as early as the 1980s, when individual institutions were encouraged to develop approaches suited to their own unique missions and student clienteles (Ewell, 1994). Such a decentralized assessment process is, however, subject to several limitations. First, many institutions were lagging in implementing credible local assessment programs, which was further exacerbated in the 1990s, when fiscal realities allowed neither states nor institutions to afford assessment as an add-on. In addition, institution-centered reports were often rich but diffuse, which failed to communicate the assessment results effectively to outside audiences. Last and probably most important, the diversity of measures employed by individual institutions rarely allowed policymakers the opportunity to compare institutions or to examine overall system performance.

As a result of increasing pressures for familiar, easy-to-read public accountability reports, states started to use an array of easily quantifiable statistical performance indicators in the 1990s. What measures should federal or state agencies or researchers use to evaluate institutional performance? Because of the multidimensionality of higher education, it has been difficult to come up with meaningful and easily calculated measures. Discussions of outcome measures as they relate to accountability have generally focused on three types: learning, employment and completion (or, more generally, measures of progression). Each of these types of measures has advantages and disadvantages, although all practical systems so far rely on measures of progression, primarily variants of completion rates.

Learning outcomes. Certainly a central mission of colleges is to teach general academic and more specific occupational skills. The burgeoning assessment activity in the K-12 sector is based on gains in standardized test scores, but such tests are much more difficult for higher education. Defining and assessing a common set of skills and knowledge for a college graduate is orders of magnitude more difficult and controversial than the similar exercise for 7th graders or even high school graduates.

There are now several assessments of general skills, the types of skills that are not necessarily taught in a specific course but are presumably learned in college. The College Learning Assessment (CLA) has been widely discussed recently and is the basis of *Academically Adrift* (Arum & Roksa, 2010), a book that argues that many students learn very few of these general skills in college. The

CLA is an essay-based assessment that measures skills such as the ability of students to use evidence to make a written argument. Other assessments, such as ACT's College Assessment of Academic Proficiency, are multiple-choice exams, and the results are highly correlated with those of the CLA. ACT's WorkKeys assessment measures more work-related skills. The National Center for Public Policy in Higher Education recommended that states use the CLA to assess learning in four-year colleges and WorkKeys to do so in two-year colleges, although their recommendations are based on using a sample throughout the state, not at the institutional level. These exams suggest that in principle it might be possible to use a test of general skills to assess learning in college, but these assessments are now commercial products that a subset of colleges pay to have administered. Mandating colleges to administer the CLA or a similar test seems highly unlikely.

There also exist assessments of learning for specific courses, especially large introductory courses, as well as for fields of study and occupational areas (often through certification exams). These could be used to measure progress at the program or course level, and this is worth pursing as part of an overall effort to improve teaching and program performance, but it would be extremely difficult to aggregate results from these types of assessments to derive an institutional measure. Furthermore, colleges now use many different versions of these types of assessments, so even for courses or within majors, there are no universally used measures that could be used to compare, in a consistent way, large numbers of colleges.[1]

Finally, the accreditation agencies are increasingly concerned with learning outcomes assessments (LOAs), but the accreditors expect colleges to define their own learning objectives and develop relevant assessments. This would allow a comparison of colleges only in the most general way. Thus progress is being made on developing learning outcomes assessments, but at this point and in the foreseeable future they are not practical as measures of institutional performance against common criteria.

Wages and employment. While assessments may measure some skills conceptualized by the test developers, it is easy to argue that they will miss many dimensions of the desired outcomes of college. Although educators emphasize that college is much more than an opportunity to learn jobs-related skills, preparation for employment is certainly a fundamental goal of a college education. Thus a measure of the wage and employment effects of college could capture a wide range of attributes, including academic and occupational skills as well as affective characteristics, such as persistence, that are more difficult to measure in an assessment but are nevertheless possible outcomes of a college education. Thus many researchers have used wages or employment to measure college achievement (e.g., Brewer, Eide, & Ehrenberg, 1999; Dale & Krueger, 2002; Cunha

[1] See Cunha and Miller (2009) for a more detailed discussion of the problems of using test scores to assess institutional performance.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

& Miller, 2009). Cunha and Miller explicitly discuss the advantages of wages as a proxy for higher education quality: Wages signal productivity, which is the byproduct of skills learned in schools; wages also reflect both general skills and specific skills acquired in college.

But there are limitations to the use of wages and employment in measures of institutional performance. First, wages cannot represent the entire set of college gains, especially those that can hardly be measured in economic terms, such as civic-mindedness. Yet even for productivity-related skills, wages may not serve as a perfect proxy. Labor market positions with lower wages do not necessarily mean that they require less knowledge; instead, individuals working in these positions may be highly motivated, such as in the nonprofit sector. In addition, wages are also largely influenced by factors that are beyond the control of individuals, such as the general economic and political condition. Further, wages are observed only for those who work. Therefore, failure to control for the probability of labor market participation may bias the assessment. For example, if an institution poorly prepares its students for the labor market and is therefore subject to a low employment rate, graduates from this college who are able to find a job may have some intrinsic attributes that also lead to higher wages. Related to this problem is a long list of individual characteristics, such as household income, parental involvement and parental education that can potentially influence both individual wages and the college of attendance.

Moreover, students at many institutions go on to further education after completing (or even not completing) a credential. How should the contribution to employment of an undergraduate college be measured if the student goes on to graduate school and secures employment based on a graduate degree? Calculating the employment effectiveness of a four-year or two-year college based only on those who are employed without subsequent education is subject to obvious biases.
The U.S. Department of Education recently created a Gainful Employment regulation based on measurement of the earnings three years after graduation of the graduates of specific occupational programs, suggesting that measurement of earnings, not only at the institutional but at the program level, would be possible. The reporting will use very complete data based on income tax records. This process avoids the problem of subsequent education by limiting the regulation to certificates and occupationally specific programs at for-profit institutions. These are the types of college programs that are most likely to lead directly to work.

One of the most interesting things about the Gainful Employment regulation is its use of federal tax data to measure wages. These are very complete data and are not limited geographically. But they have been used only under very restricted conditions. Without them the best data for this purpose are Unemployment Insurance (UI) data. These data can and have been matched to higher education administrative data to relate characteristics of college education to employment and earnings. Subject to all of the problems discussed above, this is potentially a rich source of data for measuring

the relationship between education and employment. However, in addition to the problems of using employment to measure college performance, UI has several limitations: Its availability and the quality of the match to higher education data varies widely among states; in many cases, data are limited to one state, a particularly serious problem when judging outcomes for more selective colleges that serve a national market; the self-employed and some public sector employees are not covered; in most states the data do not include occupation; and the employed data usually consist only of quarterly earnings, not differentiating between hours or days worked and wages. Some of these problems are more serious than others, and some progress can be and is being made. Some states are linking their data to allow tracking of students into other (usually contiguous) states. Nevertheless, at this time, conceptual, methodological and practical problems prevent the development of a consistent employment measure, especially a high-stakes measure, of college performance.

Completion and measures of progression. Assessments of learning and employment outcomes are measures of the actual content of a college education, and while progress is being made in developing these measures, so far there are no reliable measures that can be used consistently to compare a wide range of colleges. As a result, most state governments have relied on measures of progression through college, primarily degree or certificate completion, to assess college performance. Among all the different proxies used by state governments, the most frequently noted have been graduation/completion rates. In a report from 1994 on state-level performance indicators in 10 states, the author found that graduation was the only type of indicator used by all of the states (Ruppert, 1994).[2] Such congruence, as noted by Ewell (1994b, p. 154), was due more to circumstances than consensus: There did not exist a reliable single index of higher education quality. Among the small pool of available state-level data, very few are easily quantifiable or enjoy consensus about importance.

Recent nongovernmental efforts, such as the Voluntary System of Accountability (a project of four-year colleges) and the Voluntary Framework of Accountability (a project of two-year colleges), refer to learning and employment assessments but rely on progression measures for consistent and comparable data.

[2] Ewell (1994b, pp. 156–157) classified indicators used by states into eight major categories: Instructional Inputs (e.g., number and performance of remedial students, SAT scores), Use of Resources (e.g., student/faculty ratios, faculty workload, class size by level), Instructional Outcomes (e.g., graduation rates, student perceptions of quality, placement of graduates, major field test scores), Efficiency (e.g., credits per faculty produced), Condition of the Asset (e.g., total revenues per student, number/proportion of programs accredited), Diversity (e.g., graduation rates by ethnicity, faculty diversity), K-12 Linkages (e.g., performance of transfers at senior institutions), Relation to State Needs (e.g., graduates in science, engineering, etc.).

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

There are several advantages to using graduation rates as a measure of institutional performance. First, this information is of great interest to institutions, students and policymakers. Most students enter college planning to graduate (although this is not always the case, as we will discuss below). Colleges presumably design coherent programs that lead to graduation, and research does suggest that there are so-called "sheepskin" effects—students earn more if they complete a degree than they would if they completed an equivalent number of credits without a degree.

Second, and perhaps most important, completion and progression data are widely available, and analysts are more likely to be able to measure them consistently for a wide range of institutions than measures of learning or employment. The U.S. Department of Education publishes graduation rates for all Title IV-eligible institutions as mandated by the Federal Student Right-to-Know (SRK) and Campus Security Act.[3] These are cohort completion rates for first-time, full-time students, tracked for 150 and 200 percent of the normal time it would take a full-time student to complete a degree at the institution (six and eight years for a four-year institution and three and four years for a community college). These rates have been criticized, especially for community college students, the majority of whom are part-time. In April 2012, the U.S. Department of Education published a plan to revise the reported completion rate. According to the plan, colleges would also report on completion rates for part-time students and would include outcome measures for students who transfer in to a community college, rather than reporting only on first-time college students. http://www.ed.gov/edblogs/ous/files/2012/03/Action-Plan-for-Improving-Measures-of-Postsecondary-Student-Success-FINAL2.pdf. In addition to the SRK rates, detailed progression and completion measures at the institutional level can be easily calculated within states using state longitudinal data sets. These data sets are being improved and linked to other data sets in many states, and their use is growing rapidly.[4]

The increased use of state longitudinal data has also promoted the use of intermediate measures of student progression, such as completion of the developmental education sequence, initial enrollment in college-level courses, enrollment in programs of study, attainment of credit milestones and various measures of retention. These measures are particularly important in efforts to understand or diagnose the causes of poor institutional performance.

Although progression and completion measures can provide important insights into the performance of institutions, they have a major disadvantage—there is no consistent measure of the substantive

[3] Title IV eligibility means that students enrolled in the institution are eligible to receive federal financial aid.

[4] See Jenkins (2006) and Cunha and Miller (2009) for examples of studies that compare college performance using unit record state data sets.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

content of a degree, certificate or level of credit accumulation. Credentials represent different content at different colleges. Moreover, high-stakes accountability measures based on progression would create incentives to lower standards. For these reasons, it is important to continue to work on more direct measures of skills and effects of college.

2.2 History of College Rankings

.*S. News and World Report (USNWR)* issued the first list of its now-famous college rankings in 1983. The system ranked four-year colleges within Carnegie Classifications (i.e., national universities, national liberal arts colleges, regional universities, etc.). The methodology has changed over the years, but most of the elements used in the calculations are measures of inputs (selectivity, faculty and financial resources) rather than outcomes. The latest rankings give 20 percent weight to absolute graduation and retention rates, and for national colleges and universities 7.5 percent weight is given to graduation rates adjusted for incoming student characteristics. Thus these cannot be taken as relative measures of college performance, and the *USNWR* and other similar rankings do not include community colleges.

Consistent, comparable and public measures of college performance are relatively recent. For a college's students to be eligible for federal financial aid, colleges must report information to the Integrated Postsecondary Education Data System (IPEDS). In 1990, Congress passed the Federal Student Right-to-Know (SRK) and Campus Security Act, which mandated that colleges and universities report to IPEDS their graduation rates for fall semester cohorts of first-time, full-time students, thus enabling easy comparisons across institutions. But these rates became available only in the late 1990s. During the first decade of the 21st century, there have been some minor improvements in these so-called SRK graduation rates, including requiring eight-year rates for four-year colleges and four-year rates for community colleges, and some improvements in information on transfers.

Yet researchers (e.g., Archibald & Feldman, 2008; Astin, 1997; Calcagno, Bailey, Jenkins, Kienzl, & Leinbach, 2006, 2008; Kelchen & Harris, 2011; Mortenson, 1997; Scott, Bailey, & Kienzl, 2006) have long expressed concerns about using raw graduation rates as an objective standard for comparisons among institutions, for doing so fails to take into account student characteristics and available resource levels that have substantial impact on the likelihood of college graduation but are beyond the control of the colleges themselves. For example, competitive colleges may or may not contribute more to graduating their students compared with less competitive ones, yet the former can significantly outperform the latter in raw graduation rates simply because they have attracted students with stronger academic backgrounds.

In order to address the shortcomings of raw graduation rates, researchers have proposed developing models to adjust simple graduation rates for student characteristics and institutional resource levels. Numerous studies have been conducted to explore potential inputs that need to be taken into account as well as the appropriate techniques to control for them. In general, one can learn three main lessons from existing studies on input-adjusted graduation rates in higher education. The first is the importance of using input-adjusted metrics rather than raw graduation rates for college evaluation or comparisons across institutions, where the choice of "inputs" to be included depends on the purpose of the analysis. The second is that different techniques to measure graduation rate performance and the use of different outcome measures can lead to significantly different results and college rankings. The third is that the measurement of performance of community colleges is particularly difficult and complicated.

3. INPUT-ADJUSTED COLLEGE EVALUATION

3.1 The Role of Colleges: Why Do We Need to Adjust Inputs?

At this point, completion and progression are the best measures of college performance available, but absolute measures of progression cannot be used to compare performance among colleges. Comparing the performance of an individual over time is somewhat more defensible, but still not ideal. While graduation rates are important measures of institutional performance, they are products of the joint inputs from individuals, institutions and local governments and, more important, many of these inputs are beyond the control of the institution. Therefore, many researchers (e.g., Archibald & Feldman, 2008; Kelchen & Harris, 2011, Scott et al., 2006) have raised concerns about using raw graduation rates for college evaluation based on three major arguments.

Above all, practical outcome indicators such as raw graduation rates seldom tell policymakers, states and the public directly what they want to know. This is largely because instructional outcomes are the joint product of entering student characteristics, resource inputs and instructional processes. Using graduation rates, for example, policymakers can determine which institutions are graduating more of their students; however, what they will not know is whether this outcome is due to better-prepared entering students and available resources, or due to college practices that influence students' outcomes.

In addition, it is unfair to evaluate an institution without accounting for the characteristics of the students who enter it and the resources available to it. Because of different admission criteria, the characteristics of students when they first reach campus can be substantially different across institutions. More-selective colleges enroll more-accomplished students who start with "a distinct advantage in terms of the academic ability, educational aspirations, level and clarity of career ambition, and family resources" (Pascarella & Terenzini, 1991, p. 347), while open-access colleges,

with their clear mandate of access to underrepresented communities, enroll a large proportion of poorly prepared students. Public colleges, compared with their private counterparts, also enroll a disproportionately larger number of nontraditional students, defined as those with at least one attrition risk factor.[5] This is particularly true for public two-year colleges, where 89 percent of the students are at least minimally nontraditional, compared with 58 percent at public four-year institutions and 50 percent at private not-for-profit four-year institutions (Choy, 2002). Failing to account for student profiles also distorts rankings based on earnings (Tracy & Waldfogel, 1997; Kreutzer & Wood, 2007). Based on this reasoning, if good colleges are defined as those that can effectively prepare their students for graduation through high-quality instructional programs and school management, rather than those that attract academically talented and wealthy students through their historically built reputations, we will need to adjust for pre-college student characteristics for any college performance measures, including graduation rates. Similar arguments have been made for the necessity of controlling for resources available to institutions if we do not want to reward colleges directly for having and using more resources, regardless of whether they use them efficiently (e.g., Archibald & Feldman, 2008; Kelchen & Harris, 2011).

Finally, the focus on graduation rates without appropriate control for student heterogeneity and resource levels tends to create false short-term incentives for action, which might bring about detrimental outcomes to both institutions and society. If the stakes associated with graduation rates are high, colleges may act to maximize the numeric value of this indicator without really changing practice or performance. For example, institutions may manipulate graduation rates by restricting access to those students who have a better chance to graduate. These practices not only will weaken the country's commitment to broad-based access to high-quality higher education opportunities, but also may bring about undesirable outcomes to universities. As pointed out by Mause (2009), excessive competition for better students may lead to unproductive use of resources, as colleges tend to expend a large amount of resources in a few categories, such as targeted scholarships and campus recreation. Given these concerns, researchers seem to have reached a consensus that reasonable and meaningful judgments about the institutional graduation rates cannot be formed until academic, financial and other institutional fixed factors are well controlled for.

3.2 Determinants of Graduation Rates

The first step toward input-adjusted institutional comparisons is to identify key factors that have an impact on institutional graduation rates. Existing studies following this approach have identified a

[5] The National Center for Education Statistics (NCES) has listed seven risk factors: 1) part-time attendance; 2) full-time work; 3) delayed postsecondary enrollment; 4) financial independence; 5) having dependents; 6) being a single parent; and 7) not possessing a high school diploma. A nontraditional student is one who has any of these characteristics (National Center of Education Statistics, 2002).

mixed set of such factors, where some are under the control of colleges and some beyond. Table 1 (see Appendix) summarizes significant determinants of institution graduation rates in key recent studies. Given the substantial difference between four-year universities and two-year colleges in terms of the mission of the institution, student population served and resources available, Table 1 presents studies that focus on four-year universities (upper panel) separately from those on two-year colleges (lower panel).

As shown in the upper panel of Table 1, researchers have developed extensive empirical literature to explore the impact of individual and institutional factors on graduation rates over the past two decades. These factors fall into one of the three major categories: 1) financial variables, which include both overall institutional resources and specific resource allocation; 2) student compositional variables, which include academic selectivity, student nontraditionality and additional composition demographics; 3) institutional variables. Most studies focus only on one or two of these categories, rather than the complete set. The remainder of this section details these three major categories as well as their subcategories one by one.

Financial variables. Difficulties in accurately measuring institutional resources are well documented in Winston (1998). Based on data availability, most researchers (e.g., Archibald & Feldman, 2008; Kelchen & Harris, 2011) use expenditures to measure institutional resources, while some others (e.g., Scott et al., 2006) combine expenditures with reliable revenue measures (i.e., average undergraduate in-state tuition). Studies that include measures for institutional resources generally identify a significantly positive impact of greater overall resources on graduation rates, highlighting the importance of adjusting for the resource levels in college evaluation. Higher in-state tuition is also a significant predictor of higher graduation rates for private colleges but not so for public ones. This is reasonable if graduation rates depend on total resources available, since tuition levels are a much better proxy for resources in a private college than a public institution that can rely on state and, in some cases, local public subsidies.

However, one major problem with overall expenditure data is that such totals include expenditures not related to instructional activities and are therefore "not particularly helpful in understanding the impact of changes on the quality of higher education" (Hansen & Stampen, 1996, p. 295). Based on this account, increasing weight is given to explicit decisions about specific allocations of resources, reflected by such indicators as average expenditures per FTE in instruction, academic support,[6]

[6] Academic support includes expenses for activities and services that support an institution's primary mission of instruction, research and public service (Calcagno et al., 2008), such as academic administration and curriculum development, libraries, audio/visual services and technology support for instruction.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

student services[7] and financial aid. Empirical studies focusing on expenditure variables (e.g., Hamrick, Schuh, & Shelley, 2004; Ryan, 2004) generally suggest a positive and significant relationship between instructional and academic support expenditures and graduation rates. For example, Scott et al. (2006) found that every additional $1,000 of instructional expenditures per student leads to a 2- to 4-percentage-point increase in graduation rates, depending on the model specification. In a similar vein, Hamrick et al. (2004) also found that provision of institutional financial aid is a statistically significant predictor of cohort graduation rates. In contrast, student service expenditures do not appear to have a positive or significant effect on degree attainment (Hamrick et al., 2004). These observations, as noted by Ryan, demonstrate that there are trade-offs in the utilization of financial resources within an institution in terms of graduation rates, and that institutions should be careful when deciding where to allocate resources.

Student compositional variables. Typical measures of college selectivity for four-year institutions are SAT scores (e.g., Mortenson, 1997; Goenner & Snaith, 2004; Cunha & Miller, 2009),[8] ranking in high school classes (e.g., Archibald & Feldman, 2008) and high school GPA (e.g., Goenner & Snaith, 2004). Incoming freshman SAT scores not only serve as a reliable proxy for students' academic capacity and attachment but also enable easy comparisons across institutions. These measures for academic ability are consistent, significant predictors of graduation rates across studies. Some researchers also use the difference between the 75th and 25th percentiles to capture the variability within an institution. Scott et al. (2006), for example, found that greater dispersion of incoming SAT scores leads to significantly lower graduation rates.

While the National Center for Education Statistics (NCES) has listed seven possible characteristics of nontraditional students (see footnote 5), studies that have controlled for this factor often use student age (e.g., Porter, 2000) and part-time enrollment (e.g., Mortenson, 1997) as proxy measures. Some recent studies have also emphasized the importance of controlling for residential status (e.g., Scott et al., 2006) in college performance assessment based on the concern that nonresidential students have less opportunity to develop the type of engagement that underlies the concept of social attachment, which the theory on college completion consistently stresses as an important factor (Pascarella & Terenzini, 1991, 2005). Indeed, Scott et al. (2006) identified a significant impact from the percentage of commuter undergraduate students on institutional graduation rates even after controlling for the percentage of full-time students and the average undergraduate age, with

[7] Student services include expenses for admissions, registrar activities and activities whose primary purpose is to contribute to students' emotional and physical well-being (Calcagno et al., 2008).

[8] Mortenson (1997) and Goenner and Snaith (2004) used institution-level data, which includes information on the average SAT scores of the students, while Cunha and Miller (2009) used individual-level data, where individual SAT scores/100 is included in the estimation equation.

the latter two being significant predictors as well. This finding highlights the importance of controlling for commuter status in college performance evaluation. However, the trade-off is that this variable is not totally dependent on individual choice but might partly reflect institutional policies on social integration.

In addition to selectivity and nontraditionality, researchers have incorporated other student characteristics into their analyses, such as percentage of females (e.g., Goenner & Snaith, 2004), percentage of minorities (e.g., Scott et al., 2006) and percentage of foreign students (e.g., Scott et al., 2006). All of these variables are generally significant, where colleges with higher percentages of female students, lower percentages of Native American students and lower percentages of foreign students are associated with higher graduation rates.

Institutional variables. Researchers have included a variety of institutional characteristics in analyses of determinants or correlates of graduation rates. Many analyses use the following four variables: 1) whether the college is private or public; 2) religiosity of the college; 3) the college's location in an urban, suburban or rural area; and 4) college program emphasis. Researchers (e.g., Pascarella & Terenzini, 1991; Mortenson, 1997; Scott et al., 2006) found that private and religious institutions (e.g., Catholic colleges) tend to have higher graduation rates, and such advantage persists even after controlling for student selectivity and resource availability. However, the impacts of institution location are less consistent in the literature. Hamrick et al. (2004) and Scott et al. (2006) have identified a positive impact of more urbanized location on graduation rates while Goenner and Snaith (2004), in contrast, have found that graduation rates are higher in schools not located in urban areas.

Some studies (e.g., Astin, Tsui, & Avalos, 1996; Mortenson, 1997) have also explored potential variation between majors in degree requirements and correspondingly different graduation rates across institutions with disproportionate program emphasis. For example, Mortenson (1997) found that institutions with a lower percentage of students studying in engineering fields (e.g., institutes of technology) are associated with higher graduation rates even after controlling for other institutional characteristics. He explained that institutes of technology may be producing a different "product," a science graduate, and that a science graduate is more difficult to produce than a standard graduate. This finding also suggests that program-level analyses may be important components of any comprehensive analysis of college performance.

In contrast to fixed institutional characteristics, other characteristics are more malleable, although changing them may be difficult. These include institution size and campus housing availability. Of these variables, institution size has been the most studied, and the results are mixed. Scott et al. (2006) and Ryan (2004) found that larger enrollment has significant positive impacts on graduation

rates, while Porter (2000), in contrast, found a negative relationship between size and graduation rates. One possible explanation for such contradictory results might be the differential effects of institution size on degree attainment by student attributes. For example, Alfonso (2006) found that while Hispanic persons who enroll in large community colleges are less likely to earn an associate degree, this effect does not hold for other ethnic groups. Studies have also identified a positive relationship between campus housing availability and graduation rates, explaining that residential campuses may promote social integration, which has a positive impact on student outcomes (Pascarella & Terenzini, 1991; Porter, 2000).

3.3 Input-Adjusted Comparisons: What Inputs Do We Need to Adjust?

Input adjustment involves carrying out an analysis of outcomes that controls for demographic and institutional characteristics so that reasonable comparisons can be made among the outcomes of different institutions. The most common approach is to regress graduation rates on a set of variables and examine the residuals. Positive residuals suggest that colleges are doing better than would be predicted by the variables included in the regression. Lowest-ranked colleges are those with the largest negative residuals. We will describe this method and some alternatives later, but in this section we discuss the variables that should be included in such an analysis. This is a complex question that depends fundamentally on the purpose of the analysis and the target audience.

If the aim of a study is to understand how to improve student outcomes by identifying key factors associated with graduation rates, then it makes sense to include a comprehensive set of variables, including ones that are not under the control of colleges and others that are. In this type of study, analysts are interested in the coefficients of these variables. Examining the residuals could lead to further investigation, perhaps through fieldwork, that would try to search for causes of remaining differences in performance, even after measurable institutional and compositional characteristics have been taken into account.[9]

However, if the goal of a study is to compare or assess institutional performance, then some variables should be included while the effects of some variables should be left in the residual. All analyses of this type control for demographic variables, since the fundamental goal of the analysis is to partition the causes of outcomes between the effects of students' entering characteristics and the effects of whatever the college is doing to educate those students. (As we shall see below, controlling for demographic characteristics is not as straightforward as it may first seem.) Perhaps the most important variables measure entering students' academic skills, and these include average and variation in SAT, ACT or assessment (if they are available) scores. The admission ratio is another indication of the skills of entering students. Socioeconomic status (SES) is also an important

[9] See Jenkins (2006) for an example of this approach.

determinant of student success in college, so most studies include some measure of it. The most common is the percentage of students receiving Pell grants or some other form of financial aid. This is not a very good proxy for SES, especially for community colleges, where tuition may be low and many students attend part time and may not be eligible for financial aid. If data on individual students can be used, then geo-coding student home addresses and using Census data for small geographic areas is one increasingly used approach (Crosta, Leinbach, & Jenkins, 2006). Nontraditional status—delayed postsecondary enrollment, part-time attendance and interrupted attendance, and full-time work—also influences a student's probability of graduating, so some or all of these are usually included. Other characteristics of nontraditional students, such as nonresidential status, financial independence, having dependents, being a single parent and not possessing a high school diploma, also influence the probability of completion. Previous literature has also identified a series of other student composition characteristics that are related to institutional graduation rates, including gender, race, ethnicity and immigrant status (see Table 1 in the Appendix for variables included in various studies).

Which institutional variables should be included? Clearly some institutional factors should be left in the residual. For example, if a college uses a particular pedagogic strategy or curricular strategy, it would not make sense to adjust outcomes for that since a pedagogic strategy may lead to better outcomes. Controlling for a successful practice will reduce the residual and lower a college's ranking even though it has found a successful strategy. But note that this is exactly the type of practice that we would like to include if the purpose is to explain variation in outcomes.

Educational expenditures are an important determinant of outcomes and a fundamental component of the value-added model. These can be included as a variable in the regression, although Kelchen and Harris (2011) argue that it makes more sense to divide the outcome variable by the cost to derive a measure that is a unit of value added—for example, per $1,000 of expenditures. One significant problem with measures of cost, especially in four-year colleges, is that cost data often do not differentiate between instructional and unfunded research costs.

As Kelchen and Harris (2011) argue, the expenditures or costs included in a ranking system also depend on the target audience. Parents are concerned with the cost of a unit of outcomes to them—tuition minus institutional financial aid. This calculation will give a different ranking from one aimed at state policymakers, who may be interested in the cost to the state. Less parochial policymakers might want to take into account the total expenditures per unit of outcomes. Kelchen and Harris (2011) show that rankings are different for models using tuition minus institutional aid and those using total expenditures.

Other than expenditures, it is not clear why any institutional variable should be taken into account. If it is much more difficult to educate students in a particular type of college, say a Carnegie research institution or a college in an urban area, why should this be taken into account in rankings of student outcomes, unless there is an independent value in the feature under consideration (for example, if the analyst or consumer wants to know the best religious or urban college)? The *USNWR* rankings separate research universities from liberal arts colleges, and most research controls for Carnegie Classification, but a parent or student may want help in deciding between Swarthmore and Penn, and these rankings cannot help in that decision. While total expenditures should be taken into account, the distribution of expenditures among functions in the college should clearly be left for the residual, since this is a fundamental variable that colleges can use to improve outcomes with a given level of resources.

If the question concerns the efficiency—how well a college does with a given set of fixed resources and institutional characteristics—then taking account of institutional factors that cannot be changed, or that can be changed only with difficulty, is a reasonable approach, and this is the most common practice. These characteristics include urbanicity of the institution, whether it is private or public, its religiosity, the Carnegie Classification and whether the college is technical or comprehensive. Analysts separate two- and four-year colleges, although in practice, in most cases, two-year colleges are ignored. As we shall see, in the sub-baccalaureate arena, separating colleges that give associate degrees primarily from those that specialize in certificates is also important.

Although taking account of fixed institutional characteristics is a common practice, many apparently fixed characteristics can be changed. Institutions can grow or shrink or create quasi-independent campuses; religious colleges have separated themselves from their founding churches; colleges have changed Carnegie Classifications, added dorms, opened branch campuses (thereby changing their urbanicity) and changed their mix of programs. Community colleges have added bachelor's degree programs (and in some cases essentially transformed themselves into four-year colleges), and large research universities can include smaller liberal arts "colleges" within their overall structure.

Controlling for demographic characteristics can also create distortions. Since most value-added analyses are based on IPEDS, the calculations must be based on aggregate data. Analysts include the average SAT score because students who enter with weak academic skills are less likely to graduate (or achieve other outcomes). As we have said, adjusting for differences in outcomes related to the characteristics of entering students is the fundamental goal of value-added projects. But in aggregate analyses, these variables represent two different effects. First, the students with weak academic preparation will have a lower probability of graduating. This is an effect that should be taken into account in an input-adjustment system. But because of peer effects, high concentrations

of students with weak academic skills may have an influence on the success of students who enter with average or strong academic skills. This is an effect that should be left in the residual.

This problem can be avoided by using unit record data linked to institutional data. This allows analysts to control for individual student characteristics but also include the same characteristic as an institutional feature. Using individual data is not possible at this time using national data, although developments with the National Clearinghouse data may make this possible in the future. Certainly any state developing an internal accountability system should use individual, not aggregate, data.

4. METHODOLOGIES FOR MEASURING GRADUATION RATE PERFORMANCE

Provided with the rich set of factors in the literature that influence graduation rates, researchers have explored ways to adjust relevant inputs for raw graduation rates. We describe four different strategies for making appropriate comparisons among college outcomes: 1) selection of appropriate comparison colleges; 2) value-added analysis based on regression residuals; 3) Bayesian Model Averaging; and 4) a production frontier approach based on Data Envelopment Analysis.

4.1 Selecting Comparison Colleges

Comparing the outcomes for Columbia College and Bronx Community College) is clearly inappropriate, but how should comparison colleges be chosen? One approach is to identify key institutional characteristics, such as urbanicity, the SAT scores of incoming students or institution size, and restrict comparisons within a selected group of colleges (e.g., Muraskin & Lee, 2004). This is easy to do if the analyst focuses on a very small number of variables (comparing a college to all urban colleges or all urban colleges with enrollments over 10,000), but it becomes more complicated when several variables are used. If the method requires close similarity on several variables, then the number of potential comparison colleges will be quickly reduced. Focusing on a small number of variables will leave a great deal of heterogeneity among potential comparison colleges (Calcagno et al., 2008).

College Results Online,[10] a service of The Education Trust, has developed a methodology for using several variables to calculate a comparison score for any two four-year colleges in its database. The outcome variable is the SRK six-year graduation rate. The method first regresses the graduation rate on 10 variables, including student and institutional characteristics. The coefficients are used to

[10] http://www.collegeresults.org/aboutthedata.aspx

derive weights based on which characteristics are most closely related to graduation rates; the median SAT or ACT score is four times more important than the next most important variable (public or private). These weights add up to 1,000. If a comparison college is identical to the college under consideration with respect to a variable, then the comparison college is given the full value of the weight. As the values for the variables diverge, the college is given some percentage of the weight. Two colleges identical on all 10 variables have a comparison score of 1,000. Some other adjustments or filters are made, but basically a college can then choose the 1, 10, 20 or any number of colleges with the highest comparison scores. Colleges can then directly compare their graduation rates with other similar colleges.

The usefulness of these comparisons depends on which variables are chosen to include in the initial regression. For example, the analysis includes whether the college is public or private, the Carnegie Classification and the total enrollment. We have questioned the inclusion of these variables at least for some purposes. Costs enter only as a filter, eliminating potential comparison colleges with very different costs per FTE. Thus the method yields comparisons of reasonably similar colleges, although the results might be of more interest to a college president who is interested in comparing her college to similar ones than to a parent or student looking for a college that will teach them the most or give them the highest probability of graduating. Moreover, Kelchen and Harris (2011) argue, a comparison-college approach does not allow analysts to compare outcomes for dissimilar colleges.

4.2 Measuring College Value Added

The most common method to evaluate an institution's contribution to student outcomes is to calculate the residual from a regression equation that controls for student and college characteristics. Institutions with a positive residual are viewed as adding "value" to student outcomes. The term "value-added" is something of a misnomer in this context. Value-added analysis for K-12 institutions analyzes gains in test scores, while the typical analysis for higher education compares graduation rates taking account of input characteristics. *USNWR* refers to this method as "graduation performance."

This is an intuitively appealing input-output model that enables evaluation of institutional practice, policy and management after controlling for student input, the missions of the institutions and the constraints they face. Kelchen and Harris (2011) estimated college performance in terms of graduation rates based on the value-added approach and compared their results with three popular U.S. college ranking systems: *U.S. News and World Report*, *Washington Monthly* and *Barron's*. The comparisons between the value-added ranking and popular-college ranking revealed a low correlation, highlighting the fact that failure to control for student and institutional characteristics can result in meaningless and even misleading information for students, parents and policymakers.

Although regression adjustments can take account of measured characteristics, they do not adjust for any unmeasured factors that both determine whether a student chooses a college and influence outcomes. Cunha and Miller (2009) used information on student applications and acceptances to adjust for these unmeasured factors. Following a method developed by Dale and Krueger (2002), they in effect compared output for groups of students who graduated from high school in Texas in 1998 and 1999 and enrolled in a four-year Texas college the following year and who applied to and were accepted at the same set of colleges. The method holds that the admissions process takes account of a wide variety of factors not captured by typically measured characteristics. The decision to apply to a particular set of schools reveals something about the students as well. Thus by adding to the type of regression used in the Ordinary Least Squares (OLS) value-added models dummy variables that match students with the same application and acceptance pattern, Cunha and Miller account for a significant amount of otherwise unmeasured information, thus increasing the validity of comparisons of college-level outcomes.[11] They show that rankings of the 30 public four-year colleges in Texas change significantly when going from ones based on raw graduation rates, to the OLS value-added model, to the matched analysis. They also show that using the matched analysis, rankings also change for different outcome measures—graduation, persistence and earnings.

The student-matching model does control for unmeasured characteristics, but it requires detailed information on admissions and acceptances. These data are not widely available, even in many state data sets. Also, it is primarily usable with students who are going directly from high school to college. Thus it provides information on the effectiveness of colleges in serving these students. It is not surprising that Cunha and Miller did not use it for community college students, many of whom are older and for whom the application process is often simply showing up at the college and registering.

The OLS version of the value-added approach can be implemented with publicly available data; it uses widely known and used statistical techniques and has an easily understood intuition. It does allow the use of all of the colleges in a sample and therefore means that any college can in effect be compared with any other college, even one that is dissimilar. It gains that capability by imposing a functional form on the relationship between inputs and outputs. On the other hand, the validity of the results is also subject to the validity of the functional form, and the results suffer from the biases caused by unmeasured characteristics.

[11] This analysis still does not control for whatever factors prompted the students with the same application and acceptance patterns to choose to go to different colleges among those to which they were accepted.

4.3 Bayesian Model Averaging

As the value-added approach has become increasingly popular among studies on college accountability, some researchers (e.g., Goenner & Snaith, 2004) have pointed out a potential drawback of using straightforward regression techniques in predicting graduation rates: a heavy reliance on the accuracy of the model that is used to make the prediction. These researchers argue that most of the studies predicting college graduation rates make a strong assumption that "the variables they select to form their models are those which causally explain the data" (Goenner & Snaith, 2004, p. 11) or are, namely, the "true" model, which is an untested assumption and can be far from the fact. Indeed, as presented in Section 3, current studies exploring input-adjusted college graduation rates differ widely in their variable selection and model specification. This, as noted by Goenner and Snaith (2004), can be largely due to the uncertainty of appropriate measures for theoretically important factors on graduation rates. While theory suggests the importance of various student and institutional factors on graduation rates, it does not suggest which operational measures of these factors should be included as control variables in the model specification. For example, while there is universal agreement on the necessity of adjusting for financial constraints when evaluating college performance, there is not universal agreement on the appropriate variable that serves as the best proxy for institutional financial constraints. Thus researchers have a variable selection problem. The choice is further complicated by the theoretical ambiguity of a variable's marginal effects on graduation rates—i.e., whether a variable still has an impact on the educational outcome and therefore should be included in the model when controlling for others. The result of this is that different researchers make different decisions on whether and how to adjust for a particular factor in college performance evaluation. However, one direct outcome of such variation in model specification is a corresponding variation in the estimated effects of coefficients. This possibility is confirmed by Porter (2000) in his analysis of the robustness of the predictions made by *USNWR*. By comparing estimation results across different model specifications, Porter demonstrated that variable selection influences predicted effects of coefficients as well as their standard errors.

In view of this unavoidable uncertainty in variable selection and the potential high risk of basing inference on the predictions from a single model, Goenner and Snaith (2004) proposed using the Bayesian Model Averaging approach (BMA), which determines estimated effects by taking a weighted average of estimates over models whose specification is supported by the data. To be specific, researchers first identify a pool of candidate variables that have been theoretically linked to graduation rates;[12] since researchers are not sure of the true model, they then estimate the model based on each possible combination of variables—i.e., 2^n different models; in the final step, researchers take a weighted average of estimates over models whose specification is supported by

[12] Goenner and Snaith chose 24 such variables, relying on Tinto (1987), Pascarella and Terenzini (1991) and Astin (1993) for theoretical support.

the data.[13] The advantage of this method, according to the authors, is that it exempts researchers from an arbitrary choice between input variables. Instead, researchers can use the data to choose which measures to include. Comparing the prediction results based on BMA to those based on model selection methods that select a single model, Goenner and Snaith (2004) demonstrated that BMA provides better out-of-sample prediction for university graduation rates, as is evident in its lower mean squared error. Their findings are similar to conclusions from other studies. Colleges with better-prepared students, younger students, more women and those located in rural areas had higher graduation rates.

Goenner & Snaith (2004) used their method to calculate the determinants of graduation rates of Carnegie Research I universities, using data from IPEDS and *USNWR*. They did not develop rankings, so presumably they would use the same method as the value-added approach—ranking by the difference between predicted and actual graduation rates. So this method is not so much a new method of ranking, but rather a different method for estimating the coefficients used to calculate the residuals.

Although the Goenner and Snaith research is close to 10 years old, this approach has not been used by other analysts. Software availability may be a problem, and it is not as intuitive as the simple value-added methodology. And while it helps the analysts choose variables that are closely related to graduation rates, it does not help them choose which variables should be taken into account when making comparisons between colleges.

4.4 Production Frontiers Approach

Though differing in model specifications, the majority of the studies on assessing the graduation performance of a college have used regression techniques that compare colleges with an average educational producer. This approach has been challenged by Archibald and Feldman (2008), who argue that regression analysis may not be the best tool to assess graduation performance. They propose using a non-parametric input-output technique to establish a production frontier by identifying the colleges with the highest graduation rates for a given combination of inputs. The technique they use is referred to as data envelopment analysis (DEA). Colleges are in effect compared with the colleges that are most similar in their combination of inputs ("neighbors"). The technique calculates a technical efficiency (TE) score (from 0 to 1), which shows the percentage of

[13] In their study of determinants of graduation rates, Goenner and Snaith (2004) calculated the posterior model probability (PMP) for each of the over 16 million models possible with their 24 variables. They chose 54 models with a PMP within a chosen percentage of the PMP of the model with the highest PMP. The PMP represents the probability that model Mk is the true model that causally explains the data when conditioning on the *k* data and assuming that one of the K models is the "true" model.

the output produced by a college compared with the output of the most efficient producer with a combination of inputs that closely matches the college's amount and mix of inputs.

This approach has some similarity to the comparison college technique used by College Results Online. Although it uses a different technique to select comparison colleges, it relies on comparison to similar colleges rather than developing a method to make more comprehensive comparisons. Proponents argue that one main advantage of DEA over regression analysis is that it does not impose any structure on the relationship between graduation rates and inputs. Using non-parametric linear programming techniques, DEA makes minimal assumptions about the functional form of the production frontier (Berg, 2010); however, the trade-off is that, unlike the regression approach, it does not provide a general relationship between output and input (so this approach is less useful for identifying the variables most strongly related to outcomes). DEA and regression analysis do not necessarily provide different sets of rankings, but they are likely to do so. This is confirmed in Archibald and Feldman (2008), who compared the results between the two types of analysis and found significant differences. The results were similar for colleges near the center of the outcomes distribution but differed for colleges at the tails of the distribution. They further argued in favor of DEA, not only because of its technical advantages over regression techniques, but also because it provides several useful measures that inform an institution about how it differs from its similar colleges that are more efficient.

Archibald and Feldman (2008) conducted their analysis on 187 "national universities." They included four "input" variables: two that measure student academic preparation, one that measures the cost per undergraduate and one that measures the percentage of the faculty who are part-time. Once again this illustrates the problem with variable selection. We would question the uses of the faculty variable for some purposes. Whatever the technique, analysts still have to make appropriate decisions about their models.

DEA is an appealing approach. It sets a high standard and allows reasonable comparisons to colleges that have similar combinations of inputs. But like the BMA approach, it is not widely used for the analysis of colleges. Software availability may be a problem, as it is with BMA. The technique for selecting comparison colleges is less intuitive, or at least less well understood, than the regression approach used in the value-added analysis.

5. GRADUATION RATES IN COMMUNITY COLLEGES: A COMPLICATED ISSUE

The large majority of efforts to measure input-adjusted college performance have focused on four-year colleges. College Results Online and most value-added research consider only these colleges. The attempts to use BMA and DEA are not only limited to four-year colleges but include only research universities within their samples.

Several factors explain this neglect. First, the IPEDS graduation rate, which is the only consistent measure for a national sample of community colleges, is particularly problematic for these colleges. Since the rate includes only full-time students, it does not include the more than 60 percent of community college students who attend part time. Also, the rate includes only students who graduate from their college of first enrollment, so a student who transfers to a four-year college without completing an associate degree or certificate is in effect counted as a dropout. IPEDS does request information on transfer, but those data are inaccurate (Medwick, 2009). Also, there are no measures of incoming student academic achievement. Studies of four-year colleges can use SAT or ACT scores, but many entering community college students have not taken these tests. And since tuition is low in many community colleges, financial aid receipt is a less accurate measure of income or SES than it is for four-year colleges. Therefore, input-adjusted outcomes analysis for community colleges is thwarted by a distorted outcome variable and an inability to control for crucial input characteristics.

But analysis of community college outcomes is important because, even with more comprehensive measures based on unit record data that follow students across institutions, these institutions have low completion rates (Calcagno, Crosta, Bailey, & Jenkins, 2007; Jenkins, 2011; Karp, 2011; Leinbach & Jenkins, 2008). According to a recent study on attainment among postsecondary students (Radford, Berkner, Wheeless, & Shepherd, 2010), fewer than 36 percent of first-time college students who enrolled in a community college in 2003–04 earned a postsecondary credential within six years. Of course, these low rates should be judged in the context of the serious barriers that confront community college students, such as family responsibilities, the need to work and deficient academic preparation. Such difficulty is further complicated by the much broader range of student needs, degree requirements and institutional missions in community colleges compared with four-year universities. In the following, we discuss three issues that need to be taken into account in interpreting graduation rates as accurate indicators of college performance.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

5.1 Degree- Versus Non-Degree-Seeking Students

Community colleges have long been recognized as open-door institutions, chartered to serve multiple student needs, including a broad range of non-degree objectives (Cohen & Brawer, 1996). Therefore, many of the students enrolled in community colleges are seeking neither degrees nor transfer to a baccalaureate institution. Using graduation rates as the desired educational outcome for these students thus reflects a misunderstanding of the diverse missions that community colleges assume, and it penalizes colleges in an unfair way as well.

Some researchers have argued that it is reasonable to set high educational aspirations even for these non-degree-seeking students, given that earning small amounts of credits without completing a certificate or associate degree has little economic value (Marcotte, Bailey, Borkoski, & Keinzl, 2005; Grubb, 2002). Other researchers, on the other hand, have argued that institutions should not be judged for choices made by students that are beyond their control (Burd, 2004; Scott et al., 2006; Pascarella & Terenzini, 1991, 2005), and that educational ambition is to a large extent one such choice, particularly for adult college students. However, the problem is that most of the existing school performance measures are based on institutional-level data that do not provide information on the educational intention of the students; even with individual-level data, researchers can only vaguely infer the educational objective of a student based on his course enrollment patterns. The absence of an effective way to accurately differentiate between degree-seeking and non-degree-seeking students has thus made college performance assessment a complex issue in community colleges. Some researchers have used course enrollment patterns, using individual-level administrative data sets, to infer whether a student is degree- or non-degree-seeking. Non-degree-seeking students might be more likely to enroll intermittently or to take noncredit courses, while degree-seeking students might be more likely to enroll continuously, to enroll in transfer courses or degree requirements, or to enroll full time. The main problem with this approach is that it uses descriptions of in-college behavior as a proxy for entering-college characteristics. College practice, such as poor advising or a disorganized schedule, may influence student behavior. Using course-taking patterns as a control may therefore disguise poor performance.

5.2 Transfer as a Successful Outcome

Transfer to a four-year university is the educational intention as well as a desirable educational outcome for a substantial number of students initially enrolled in community colleges (Ehrenberg & Smith, 2004; Leigh & Gill, 2003; Rouse, 1995). Yet, as we have pointed out, most existing college accountability systems count transfer out of school simply as "dropout." This might make school evaluations particularly difficult in community colleges, given the large number of students who leave community colleges prior to earning a credential and subsequently enroll in another intuition (Calcagno et al., 2006). According to a recent study on degree completion in community colleges

(Jenkins, 2011), about 15 percent of first-time students transferred to a four-year institution without having first earned any community college credential. This particular type of transfer, from a two-year college to a four-year university, is in line with the mission of community colleges. Neglecting it not only makes colleges look worse but also unfairly penalizes colleges with higher proportions of students transferring to four-year institutions.

In addition, there can be "swirling" between a community college and a nearby four-year university, in which students enrolled in the four-year university might take one or two classes in the community college. This may largely lower the graduation rates in community colleges that are close to or in partnership with a four-year university if these one-time course enrollees are all counted as dropouts. However, it is also inappropriate to count these students as "transfer to a four-year," since they are in fact four-year university students in the first place. Based on this account, it is important that state governments establish data systems that track individual students' course enrollment patterns and academic outcomes across institutions. This will allow researchers to identify "real" students in a college as well as differentiate "transfer to four-year" from "dropout from college." Community college graduation rates would about double if a measure were used that tracked students across institutions and that followed them for six rather than three years (Medwick, 2009).

5.3 Certificates Versus Associate Degrees

The IPEDS graduation rate combines data on associate degrees and certificates, but these are fundamentally different awards. Certificates take much less time to complete and are often focused on specific occupational credentials. The students attracted to an occupational certificate program and those attracted to a community college transfer program undoubtedly differ in their goals and orientation toward college. Indeed, of the 50 community colleges with the highest graduation rates, 38 did not confer a single associate degree—all credentials were certificates (Bailey, 2011). Likewise, multivariate studies on graduation rates in community colleges have consistently found that colleges with a greater emphasis on certificate programs rather than associate degree programs, such as technical colleges, are associated with higher graduation rates (e.g., Alfonso, Bailey, & Scott, 2005; Calcagno et al., 2006; Scott-Clayton & Weiss, 2011). Scott-Clayton & Weiss, for example, used administrative data from Washington state to compare the outcomes of young, career-technical students across community colleges and technical colleges; they found that technical schools have significantly higher certificate completion rates after three years but show no apparent difference in terms of associate degree completion. Calcagno et al. (2006) found that technical colleges, which focus more on certificates than community colleges, and colleges that award more certificates than associate degrees have higher graduation rates. The problem with using degree ratios is that it disguises any differences in completion rates for the two types of degrees. In any case, this finding is not surprising since it takes less time to complete a certificate and since certificate students are

likely to have more focused goals.[14] Clearly, combining associate degrees and certificates distorts graduation rates as measures of performance.

Despite these problems there have been a small number of attempts to compare community college outcomes. Calcagno et al. (2008) carried out a regression-based analysis of the SRK graduation rates using demographic and institutional variables available from IPEDS. Although the analysis did calculate residuals, these authors did not report a ranking.

Carey (2007) has carried out the most ambitious, and controversial, community college ranking project. He relied primarily on results of the Community College Survey of Student Engagement (CCSSE). CCSSE collects information on the types of experiences students have in community colleges, including information on active and collaborative learning, student effort, academic challenge, faculty-student interaction and support for learners. These factors account for 85 percent of the rankings, and the IPEDS graduation rate accounts for 15 percent. Carey argues that research has shown a relationship between these factors as measured by CCSSE and student outcomes such as grades, credit attainment and graduation. The effort to measure and compare community college outcomes is still at a very rudimentary stage.

6. CONCLUSIONS AND RECOMMENDATIONS

The focus on college outcomes is only going to increase. Whether the purpose is for general accountability, outcome-based funding, consumer choice or research, adjusting outcomes for student characteristics and some institutional inputs is crucial. Ranking colleges or judging institutional performance using unadjusted outcomes is meaningless at best and is likely to lead to negative consequences. Researchers over the past decade have made some first steps in analyzing college outcomes and adjusting results for inputs. Most of this research has used the IPEDS graduation rate, controlling for aggregate institutional variables. By far the most common approach is to examine the residuals of a graduation rate regression, although other techniques for calculating coefficients and choosing comparison colleges have also been used. While there is consensus that student characteristics, particularly measures of academic achievement, must be taken into account, there is no consensus on which institutional variables should be included or how to adjust for costs. Very little research has been carried out on input-adjusted outcomes for community colleges. Our review suggests several conclusions and recommendations.

[14] According to the IPEDS instructions, colleges should track certificate students for 150 percent of the normal time it would take to complete a certificate—13.5 months for a nine-month certificate. We suspect that colleges simply use the three-year tracking period for all credentials, therefore biasing graduation rates in favor of certificates.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

6.1 Graduation Rates

Given the availability of data, using graduation rates is a perfectly reasonable approach. They are available for almost all colleges, and the IPEDS rates are measured using a consistent approach.[15] Nevertheless, basing accountability or any assessment of effectiveness on these rates is far from ideal. There are several serious problems.

First, without controlling for the quality, graduation rates (or any progression measure) can be increased without changing underlying performance. This is why progress needs to be made on using learning or employment outcomes as a measure of the quality of progression or completion measures. We are not ready to use these measures yet, but any serious analysis of college outcomes must make progress in these areas.

Second, as we have pointed out, the IPEDS rates themselves can present a distorted image of college performance. This is one area where some progress is being made, by including part-time students and improving information on transfer.

Third, some accountability systems are beginning to include intermediate measures to gauge progress short of completion (Shulock & Jenkins, 2011). Although credentials are valuable in the labor market, there is also value in accumulation of college credits, even if students do not graduate.

Fourth, researchers have expressed concerns about the appropriateness of graduation rates as the single outcome measure for school performance (e.g., Archibald & Feldman, 2008; Bolt & Roberts, 1997; Kelchen & Harris, 2011; Lavin & Hyllagard, 1996; Muffo, 1996), arguing that this does not reflect the multiple missions of higher education institutions. Given this, it makes sense to use multiple outcomes to evaluate college performance.

6.2 Control Variables

Choosing which variables to take into account or to count as inputs is a fundamental aspect of this line of research. Including a variable in the analysis will remove its positive or negative effect from the ranking or assessment system. Discussion of this fundamental issue has been surprisingly superficial in published research in the area. We lean toward being skeptical about including institutional variables other than student characteristics and overall resource level, even those that are hard to change, and suggest that analysts need to have very good arguments about why to include institutional variables. The fundamental conclusion is that the appropriate ranking and

[15] While the methodology is consistent, there are many opportunities for colleges to make their own interpretations, so the quality and meaning of these rates vary (Albright, 2010).

assessment systems will be different for different audiences. Parents and students (and different types of students), state legislators, state-level agency personnel, college presidents, accreditors or researchers looking for efficiency or the best use of society's resources may want different information and assessment systems. Also, using different outcome measures, as we suggested earlier, also allows more flexibility to address the needs of different audiences.

6.3 Research Methodology

The regression-based graduation rate analysis has been the mainstay of input adjustment. It is easy to implement, is well understood and allows the analyst to judge performance among dissimilar colleges. Critics have argued that the approach makes arbitrary decisions about variable inclusion (Bayesian Model Averaging), that it is based on strong assumptions about model specification (Data Envelopment Analysis) and that it focuses on the average college when the research should be concerned with identifying the best college for a given set of inputs (Data Envelopment Analysis). These alternatives are probably worth further exploration, but fewer researchers are familiar with these approaches, and the necessary software may be difficult to acquire. This is probably why they have not been widely used. It would be worthwhile to carry out projects that compare and analyze rankings using different approaches. It should be noted, however, that none of these methodologies solves the problem of appropriate variable selection discussed above.

6.4 Programs Versus Institutions

Institutional graduation rates are a combination of the performance of different programs or "schools" within colleges; thus institutional aggregates may obscure great variation in performance at the program level. Different programs and majors have very different employment outcomes, and we have pointed out how certificate completion rates are much higher than associate degree rates in community colleges. The Department of Education's Gainful Employment program does focus on program-level analysis. Certainly when possible, more research should be conducted on program outcomes. One problem is that many students do not enter programs immediately, so it is not clear who is in an initial program cohort.

6.5 Aggregate Versus Individual Data

Research on college outcomes and associated input adjustments is fundamentally handicapped by its reliance on institutional-level data. Several of the recommendations that we have made so far, such as the use of multiple outcome measures and use of program-level analysis, cannot be carried out with currently available college graduation rates. As we pointed out earlier, individual-level data would allow analysts to differentiate between the lower probability of graduation for a part-time student, for example, and the effect that the presence of large numbers of part-time students has on outcomes for full-time students. Our earlier discussion of community colleges suggests that the use of unit record data that track students among institutions, or at least its use in deriving better

institutional-level outcomes, is particularly important for community colleges. The few studies that have used individual-level data (e.g., Bailey, Calcagno, Jenkins, Leinbach, & Kienzl, 2005) have revealed a rich set of individual-level characteristics that have a significant impact on program completion (e.g., college major, enrollment behaviors, ethnicity, SES), most of which are not available in institutional-level data.

If IPEDS continues to be the source for outcome measures, then research will have to rely on aggregate outcome measures, albeit somewhat improved measures. The creation of a federal-level unit record system for college students was prohibited by the last reauthorization of the Higher Education Act. The development of the National Clearinghouse database may offer some possibilities, but it is unlikely that that database can be the basis of a comprehensive project on input-adjusted outcomes. State data systems, perhaps supplemented with National Clearinghouse data, are the best possibility at this time. These do not allow national ranking studies, but they should be the basis of state-based accountability or ranking systems, especially in large states with many institutions. This is particularly useful for community colleges since states tend to have more community colleges than four-year institutions.

6.6 Limitations of Input-Adjusted Outcomes and Related Ranking Systems

Researchers and educators have made a great deal of progress in adjusting outcome measures for differences in inputs. Of course, rankings are still dominated by the *USNWR* and similar systems that not only fail to adjust outputs for inputs, but label inputs as outputs. Nevertheless, we are still a very long way from developing a reliable approach to adjusting outcomes. Rankings change with different methodologies, different outcomes and different control variables. Since reasonable arguments can be made for many of these differences, reasonable arguments can also be made for the different rankings. Relative outcome measures should not be taken too seriously. Small or even moderate differences have little meaning. Perhaps if a college appears near the top or bottom on many systems, then we start to have some confidence in those conclusions.

Despite the limitations that we have articulated, the research that we have discussed in this review represents a very important endeavor. Improvements in data availability, refinements in methodologies, coordinated use of qualitative methodologies to further explain quantitative outcomes, and growing interest spurred by demands for more accountability will lead to a new understanding of what we can do to improve college outcomes.

REFERENCES

Albright, B. (2010). *Suggestions for improving the IPEDS Graduation Rate Survey data collection and reporting*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010832

Alfonso, M. (2006). Hispanic educational attainment in sub-baccalaureate programs. In C. Horn, S. Flores, & G. Orfield (Eds.), *Latino educational opportunity: New directions for community colleges* (No. 133). San Francisco, CA: Jossey-Bass.

Alfonso, M., Bailey, T., & Scott, M. (2005). The educational outcomes of occupational sub-baccalaureate students: Evidence from the 1990s. *Economics of Education Review, 24*(2), 197–212.

American Association of State Colleges and Universities. (2002). *Accountability and graduation rates: Seeing the forest and the trees*. Washington, DC

Archibald, R. B., & Feldman, D. H. (2008). Graduation rates and accountability: Regressions versus production frontiers. *Research in Higher Education*, *49*(1), 80–100.

Arum, R., & Roksa, J. (2010). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Astin, A. W. (1993). *What matters in college*. San Francisco, CA: Jossey-Bass.

Astin, A. W. (1997). How "good" is your institution's retention rate? *Research in Higher Education*, *38*(6), 647–658.

Astin, A. W., Tsui, L., & Avalos, J. (1996). *Degree attainment rates at American colleges and universities: Effects of race, gender, and institutional type*. Los Angeles, CA: University of California, Los Angeles, Higher Education Research Institute.

Bailey, T. (2011, February). *Can community colleges achieve ambitious graduation goals?* Paper presented at the American Enterprise Institute conference, Washington, DC. Retrieved from http://www.aei.org/docLib/Can%20Community%20Colleges%20Achieve%20Ambitious%20Graduation%20Goals%20by%20Thomas%20Bailey.pdf

Bailey, T., Calcagno, J.C., Jenkins, D., Leinbach, D., & Kienzl, G. (2005). *The effects of institutional factors on the success of community college students*. New York: Columbia University, Teachers College, Community College Research Center.

Berg, S. V. (2010). *Water utility benchmarking: Measurement, methodologies, and performance incentives*. London, UK: International Water Association.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

Bolt, J. R., & Roberts, K. J. (1997, May). *What are performance indicators and can they be used to manage colleges? A comparative study of similar colleges in the US and the UK*. Paper presented at the Association for Institutional Research Annual Forum, Orlando, FL.

Brewer, D., Eide, E. R., & Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effect of college type on earnings. *Journal of Human Resources*, *34*(1), 104–123.

Burd, S. (2004, April 2). Graduation rates called a poor measure of colleges. *The Chronicle of Higher Education*, p. A1.

Calcagno, J. C., Bailey, T., Jenkins, D., Kienzl, G., & Leinbach, T. (2006). Is student right-to-know all you should know? An analysis of community college graduation rates. *Research in Higher Education*, *47*(5), 491–519.

Calcagno, J. C., Bailey, T., Jenkins, D., Kienzl, G., & Leinbach, T. (2008). Community college student success: What institutional characteristics make a difference? *Economics of Education Review*, *27*(6), 632–645.

Calcagno, J. C., Crosta, P., Bailey, T., & Jenkins, D. (2007). Does age of entrance affect community college completion probabilities? Evidence from a discrete-time hazard model. *Educational Evaluation and Policy Analysis*, *29*(3), 218–235.

Carey, K. (2007, September). America's best community colleges: Why they're better than some of the "best" four-year universities. *Washington Monthly*. Retrieved from http://www.washingtonmonthly.com/features/2007/0709.careyessay.html

Choy, S. (2002). *Nontraditional undergraduates: Findings from the condition of education 2002* (Report No. NCES 2002-012). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Cohen, A. M., & Brawer, F. B. (1996). *The American community college* (3rd ed.). San Francisco, CA: Jossey-Bass.

Crosta, P. M., Leinbach, T., & Jenkins, D. (with Prince, D., & Whittaker, D.). (2006). *Using Census data to classify community college students by socioeconomic status and community characteristics* (CCRC Research Tools No.1). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from: http://ccrc.tc.columbia.edu/Publication.asp?uid=430

Cunha, J. M., & Miller, D. W. (2009). Quantitative measures of achievement gains and value-added in higher education: Possibilities and limitations in the State of Texas. Retrieved from: http://faculty.nps.edu/jcunha/Cunha_Miller_College_Value_Added_Report.pdf

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.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

Dougherty, K., Natow, R., Hare, R., Jones, S., & Vega, B. (2011). *The politics of performance funding in eight states: Origins, demise, and change*. New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/Publication.asp?uid=875

Ehrenberg, R. G., & Smith, C. L. (2004). Analyzing the success of student transitions from 2-year to 4-year public institutions within a state. *Economics of Education Review, 23*(1), 11–28.

Ewell, P. T. (1994). Developing statewide performance indicators for higher education: Policy themes and variations. In S. S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators* (pp. 147–166). Denver, CO: Education Commission of the States.

Goenner, C. F., & Snaith, S. M. (2004). Accounting for model uncertainty in the prediction of university graduation rates. *Research in Higher Education*, *45*(1), 25–41.

Grubb, W. N. (2002). Learning and earning in the middle, part I: National studies of prebaccalaureate education. *Economics of Education Review*, *21*(4), 299–321.

Hamrick, F. A., Schuh, J. H., & Shelley, M. C. (2004). Predicting higher education graduation rates from institutional characteristics and resource allocation. *Education Policy Analysis Archives*, *12*(19).

Hansen, W. L., & Stampen, J. O. (1996). The financial squeeze on higher education institutions and students: Balancing quality and access in financing higher education. In D. W. Breneman, L. L. Leslie, & R. E. Anderson (Eds.), *ASHE reader on finance in higher education* (pp. 291–302). Needham Heights, MA: Simon and Schuster.

Jenkins, D. (2006). *What community college management practices are effective in promoting student success? A study of high- and low-impact institutions*. New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/Publication.asp?uid=419

Jenkins, D. (2011). *Get with the program: Accelerating community college students' entry into and completion of programs of study* (CCRC Working Paper No. 32). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/Publication.asp?UID=885

Kelchen, R., & Harris, D. (2011). *Pulling rank: The potential of value-added to address the failures of U.S. News and other popular college rankings*. Unpublished manuscript.

Karp, M. M. (2011). *Toward a new understanding of non-academic student support: Four mechanisms encouraging positive student outcomes in the community college* (CCRC Working Paper No. 28, Assessment of Evidence Series). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/Publication.asp?UID=860

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

Kreutzer, D. W., & Wood, W. C. (2007). *Value-added adjustment in undergraduate business school ranking*. Washington, DC: Heldref Publications.

Lavin, D. E., & Hyllegard, D. (1996). *Changing the odds: Open admissions and the life chances of the disadvantaged*. New Haven, CT: Yale University Press.

Leigh, D. E., & Gill, A. M. (2003). Do community colleges really divert students from earning bachelor's degrees? *Economics of Education Review*, *22*(1), 23–30.

Leinbach, D. T., & Jenkins, D. (2008). *Using longitudinal data to increase community college student success: A guide to measuring milestone and momentum point attainment* (CCRC Research Tools No. 2). New York, NY: Columbia University, Teachers College, Community College Research Center.

Marcotte, D., Bailey, T., Borkoski, C., & Kienzl, G. S. (2005). The returns to education at community colleges: Evidence from the National Education Longitudinal Survey. *Educational Evaluation and Policy Analysis*, *27*(2), 157–175.

Mause, K. (2009). Too much competition in higher education? Some conceptual remarks on the excessive-signaling hypothesis. *American Journal of Economics and Sociology*, *68*(5), 1107–1133.

Medwick, A. J. (2009). *Transfer-out reporting on the IPEDS graduation rate survey: A first look at trends and methods*. Tallahassee, FL: Association for Institutional Research and U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved from http://airweb.org/images/Medwick_Final_Report_2010.pdf

Mortenson, T. (1997). Actual versus predicted institutional graduation rates for 1100 colleges and universities. *Postsecondary Education Opportunity*, *58*, 1–24.

Muffo, J. A. (1996, May). *Lessons learned from a decade of assessment*. Paper presented at Association for Institutional Research Annual Forum, Albuquerque, NM.

Muraskin, L., & Lee, J. (2004). *Raising the graduation rates of low-income college students*. Washington, DC: Pell Institute for the Study of Opportunity in Higher Education.

National Center for Education Statistics (2002). *Findings from the condition of education 2002: Nontraditional undergraduates* (Condition of Education Report No. NCES 2002-012). Washington, DC: U.S. Department of Education.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco, CA: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). How college affects students: A third decade of research. San Francisco, CA: Jossey-Bass.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

Porter, S. R. (2000). The robustness of the "graduation rate performance" indicator used in the *U.S. News and World Report* college ranking. *International Journal of Educational Advancement*, *1*(2), 10–30.

Radford, A. W., Berkner, L., Wheeless, S. C., & Shepherd, B. (2010). *Persistence and attainment of 2003–04 beginning postsecondary students: After 6 years* (Report No. NCES 2011-151). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Rouse, C. E. (1995). Democratization or diversion? The effect of community colleges on educational attainment. *Journal of Business and Economic Statistics, 13*(2), 217–224.

Ryan, J. F. (2004). The relationship between institutional expenditures and degree attainment at baccalaureate colleges. *Research in Higher Education*, *45*(2), 97–114.

Ruppert, S. S. (1994). (Ed.). *Charting higher education accountability: A sourcebook on state-level performance indicators*. Denver, CO: Education Commission of the States.

Scott, M., Bailey, T., & Kienzl, G. (2006). Relative success? Determinants of college graduation rates in public and private colleges in the U.S. *Research in Higher Education*, *47*(3), 249–280.

Scott-Clayton, J., & Weiss, M. J. (2011). *Institutional variation in credential completion: Evidence from Washington State community and technical colleges* (CCRC Working Paper No. 33). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/Publication.asp?uid=916

Shulock, N., & Jenkins, D. (2011). *Performance incentives to improve community college completion: Learning from Washington State's Student Achievement Initiative* (A State Policy Brief). New York, NY: Columbia University, Teachers College, Community College Research Center. Retrieved from http://ccrc.tc.columbia.edu/Publication.asp?uid=880

Tinto, V. (1987). *Leaving college*. Chicago, IL: University of Chicago Press.

Tracy, J., & Waldfogel, J. (1997). The best business schools: A market-based approach. *Journal of Business*, *70*(1), 1–31.

U.S. Department of Education, (2012) Action Plan for Improving Measures of Postsecondary Student Success. Retrieved from http://www.ed.gov/edblogs/ous/files/2012/03/Action-Plan-for-Improving-Measures-of-Postsecondary-Student-Success-FINAL2.pdf

Winston, G. C. (1998). *A guide to measuring college costs* (Discussion Paper No. 46). Williamstown, MA: Williams College, Williams Project on the Economics of Higher Education.

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

APPENDIX: TABLE 1

Summary of Key Findings from Recent Studies on Graduation Rates

Authors	Title	Major Data Sets	Method	Significantly Positive Inputs and Characteristics (significant at the 0.05 level)
Four-Year Universities				
Archibald & Feldman (2008)	Graduation rates and accountability: Regressions versus production frontiers	*USNWR*, IPEDS	Regression & data envelopment analysis	Higher SAT scores; higher per student expenditures; higher percentage of students in the top 10 percent of their high school classes
Astin at al. (1996)	Degree attainment rates at American colleges and universities: Effects of race, gender, and institutional type	The Cooperative Institutional Research Program	Regression	Higher selectivity; larger enrollments in business, psychology, and social sciences; smaller enrollments in engineering; smaller enrollments overall
Cunha & Miller (2009)	Quantitative measures of achievement gains and value-added in higher education: Possibilities and limitations in the state of Texas	Individual-level database housed at the Texas Higher Education Coordinating Board offices	Regression	Higher SAT scores; female; higher household income in high school; not being at risk of not graduating from high school
Goenner & Snaith (2004)	Accounting for model uncertainty in the prediction of university graduation rates.	IPEDS, *USNWR*	Bayesian averaging method	Higher GPA and SAT scores; higher percentage of female students; lower percentage of Native American students; lower average age; not located in urban areas
Hamrick et al. (2004)	Predicting higher education graduation rates from institutional characteristic and resource allocation	IPEDS	Regression	More urbanized location; lower percentage of applicants admitted; greater instructional expenditures, library expenditures, and academic support
Mortenson (1997)	Actual versus predicted graduation rates for 1100 colleges and universities	IPEDS, *USNWR*, 1995 Directory of Postsecondary Institutions	Regression	Higher SAT scores; higher percentage of freshmen living in campus housing; lower percentage of undergraduate students enrolled part time; lower percentage of students studying in engineering fields; Catholic college
Porter (2000)	The robustness of the "graduation rate performance" indicator used in the U.S. News and World Report college ranking	IPEDS, DILTS	Regression	Higher SAT scores; higher expenditure; smaller enrollments; higher percentage of females; lower percentage of students above age 25; greater campus housing availability
Ryan (2004)	The relationship between institutional expenditures and degree attainment	IPEDS	Regression	Greater instructional and academic support expenditures; larger enrollments

Context for Success is a research and practice improvement project designed to advance the best academic thinking on postsecondary institutional outcome measures. The project was organized by HCM Strategists LLC with support from the Bill & Melinda Gates Foundation. The papers may not represent the opinions of all project participants. Readers are encouraged to consult the project website at: www.hcmstrategists.com/contextforsuccess.

Scott et al. (2006)	Relative success? Determinants of college graduation rates in public and private colleges in the U.S.	American Survey of Colleges, IPEDS	Regression	Greater institutional resources; greater instructional expenditures; higher SAT scores; higher percentage of full-time students; lower percentage of commuter students; smaller average student age; located in urban area; higher percentage of female students; lower percentage of minority students; lower percentage of foreign students; larger enrollment; religious institution
Community Colleges and Technical Schools				
Bailey et al. (2005)	The effects of institutional factors on the success of community college students	IPEDS	Regression	*Institutional level:* Smaller enrollments; not located in urban areas; low share of minority students, part-time students, and women; greater instructional expenditures *Individual level:* from higher SES households; recipients of financial aid; full-time enrollment; non-Black ethnicity; enrolled in academic major; not enrolled in remedial courses; immediate enrollment after high school graduation; continuous enrollment in college
Calcagno et al. (2008)	Community college student success: What institutional characteristics make a difference?	IPEDS, NELS:88	Regression	Smaller enrollment; smaller percentage of minority students
Calcagno et al. (2006)	Is Student Right-to-Know all you should know? An analysis of community college graduation rates	IPEDS	Group logit regression	Less urbanized location; smaller enrollments; low share of minority students, part-time students, and women; greater instructional expenditures; technical colleges; certificate-oriented colleges
Clayton & Weiss (2011)	Institutional variation in credential completion: Evidence from Washington state community and technical colleges	Administrative data from Washington state	Regression	Technical colleges (higher certificate completion rates)