Conceptualizing and Measuring Productivity in U.S. Higher Education

Christopher Mackie, Committee on National Statistics, National Academies of Sciences, Engineering, and Medicine

TIAA INSTITUTE

NACUBO

About this Research

This paper is one of five in the *TIAA Institute Higher Education Series: Understanding Academic Productivity*, an initiative undertaken in support of NACUBO's Economic Models Project. That project was launched by NACUBO with the aim to provide colleges and universities with knowledge, ideas and tools to advance the difficult structural, cultural and political changes required for moving to more sustainable economic models. Given NACUBO's goal of offering thoughtful, objective and credible scholarship on the issues at hand, the TIAA Institute was a natural partner for the project.

This paper presents a deeply-informed review of the literature by Chris Mackie, who served as the study director for the National Academy of Sciences' 2012 report, *Improving Measurement of Productivity in Higher Education*. Mackie's extensive knowledge and understanding of the questions and nuances surrounding productivity measurement is clear. It is our expectation that his review will serve as a definitive resource for senior campus leaders and scholars wishing to delve further into the wide range of topics that Mackie succinctly addresses in this work.

About the TIAA Institute

The TIAA Institute helps advance the ways individuals and institutions plan for financial security and organizational effectiveness. The Institute conducts in-depth research, provides access to a network of thought leaders and enables those it serves to anticipate trends, plan future strategies and maximize opportunities for success. To learn more, visit www.tiaainstitute.org.

About NACUBO

NACUBO, founded in 1962, is a nonprofit professional organization representing chief administrative and financial officers at more than 2,100 colleges and universities across the country. NACUBO's mission is to advance the economic viability, business practices and support for higher education institutions in fulfillment of their missions. For more information, visit www.nacubo.org.

Executive Summary

Questions about productivity in higher education and how to measure it have been explored and debated for more than 100 years, dating back at least to a Carnegie Foundation report in 1910. Although stubborn challenges persist, discussions about productivity and performance have become far more nuanced, reflecting an expanding recognition that one-dimensional measures are insufficient for most policy, operational, and consumer information purposes. The limitations of unit cost, graduation rate, time to degree, and similar metrics, while clearly essential pieces of the required information dashboard for higher education, are now better understood than ever. Efforts to accurately measure the inputs and outputs of higher education production, and to seriously address the quality dimensions of each component alongside quantities, have accelerated. The literature reviewed in this paper reflects advances made in the modeling of economic concepts applied to the sector, the development of improved approaches for measurement of these concepts, and the findings that have emerged from research in the area. Additionally, data sources continue to evolve and are now more comprehensive than ever before, giving further promise to researchers and policy makers who seek carefully considered performance assessments of higher education institutions, systems, and the sector broadly. Remaining challenges for next steps in the advancement of conceptual frameworks and for addressing key data gaps are indentified throughout.

This paper is one in a series of five focusing on productivity and new economic models for higher education.

The series offers a deeply-informed review of the literature, a two-part examination of higher education's contributions to the public good, an assessment of state-level efforts to measure productivity, and an in-depth description of a course-based tool to analyze costs.

Key Takeaways

- For purposes of measuring productivity in higher education, metrics should be constructed after goals and objectives have been identified—otherwise, administrators and policy makers will value something that is measurable rather than measuring something that is valuable.
- When attention is myopically focused on one dimension of performance—whether unit costs, graduation rates, or some other metric—the risk is heightened that goals based on that dimension will be pursued at the expense of others including, most worryingly, quality.
- The inevitable presence of difficult-to-quantify elements in productivity and performance measurement should not be used as an excuse to ignore those elements.
- In higher education, productivity improvement is seen as the most promising strategy for containing costs in the continuing effort to expand access and affordability while keeping the quality of college education in the United States at a world-class level.

- Studies indicate that, among the total benefits of higher education to society, more than half accrue as positive externalities and public goods. Failure to capture these broader benefits distorts calculations of value added and return on investment calculations, which in turn leads to misallocation of public funds.
- Weighting the different components of value associated with higher education outcomes (for example, in terms of graduates' learning versus earnings gains, or private versus public benefits) will always require subjectivity and subject matter expertise and will be driven by the specific question being asked.

1. Introduction

Conversations about higher education—its accessibility and affordability, its effectiveness in preparing graduates to engage productively in the economy and society, and its return on public investment to individuals and to society—abound. These conversations take place among students and parents, faculty and administrators, and federal, state, and local policy makers and advisors. They are complex, addressing a wide range of questions involving many separate but interrelated (and sometimes confused) concepts and issues.

This review of the economics of higher education and its measurement is structured around *economic productivity*, a concept that provides a framework for organizing components of the process through which services from the sector are delivered. At its most basic level, Griliches (1987, p. 1010) defined productivity, slightly vaguely, as "a ratio of some measure of output to some index of input use." The measure may relate a quantity of outputs to the quantity of inputs used in their production—for example, the number of degreed graduates that can be produced with some amount of labor hour inputs. Alternatively, monetary values may be used to aggregate both inputs and outputs; this has practical appeal, especially for complex production processes, since prices reflect the relative values of the multiple inputs that must be accounted for. In the monetized application, one must not only sum the cost of all inputs, but also arrive at a value of all output, which is quite simple when its price is determined in a market. However, many of the outputs of higher education, and indeed some of the inputs, are not priced in markets which, as we explore below, creates measurement challenges.

This review of the economics of higher education and its measurement is structured around *economic productivity*.

At its most basic level, Griliches defined productivity as "a ratio of some measure of output to some index of input use."

In principle, the productivity construct encompasses the essential elements that must be measured and tracked in order to assess the performance of a sector of the economy, such as higher education, or of its sub-aggregates, such as individual institutions or groups of institutions, within the sector. The most obvious output of higher education is the production of educated individuals who go on to contribute labor of varying skills and attributes to the economy and, more broadly, to contribute in an informed way to the functioning of a civically engaged community and nation. But the outputs of colleges and universities extend far beyond credentialed graduates. Research findings (that advance science, medicine, and industry), hospital services, housing for students and faculty, cultural and sports events, and other entertainment offerings to the community are also produced. Inputs include the scarce resources drawn upon and directed toward operation of the enterprise—most notably the human capital provided by skilled faculty, administrators, and other staff, but also the physical capital and intermediate services used in the production process. Students themselves, and the opportunity cost of the time they commit, may—depending on one's perspective and the questions being asked—also be viewed as a crucial input in the education production function.

An important point in measuring productivity is that **goals and objectives should be set before measures are constructed.**

Outputs and inputs can be defined, quantified, and measured in a variety of ways; appropriate specification depends on the question or questions of interest. As succinctly put by Garrett and Poole (2006), "an important point in measuring productivity is that measures should not be constructed prior to setting goals and objectives—doing so will lead administrators to value something that is measurable rather than measuring something that is valuable." In the current environment of strong policy maker and institutional interest in the performance of higher education, well-meaning stakeholders tend to use whatever data and measures are available to better understand trends and perceived problems and to push agendas. As observed by authors of the National Academy of Sciences report, *Improving Measurement of Productivity in Higher Education* (hereafter, NAS 2012): "For better or worse, some version of productivity *will* be measured. Therefore, it is crucial to develop coherent measurement tools that make the best possible use of available and potentially available data. Failure to do so will keep the door open for an ever-expanding profusion of measures, many of them unnecessarily distortive, and endless debates about measurement as opposed to productivity itself" (p. 11).

For most purposes, tracking the quantity of inputs used and outputs produced—which is typically easy—is inadequate. If institutions or departments were evaluated on the basis of the number of graduates they produced per hour of faculty time or per dollar spent, surely a "race to the bottom" in terms of quality would be triggered as class sizes were increased and expenditures slashed. Answering most questions surrounding productivity requires an assessment of inputs and outputs in terms of their value, which requires taking into account variation and changes in their *quality*. In order to value the total output of the education sector, the quantity of graduates must in a sense be multiplied by a measure of the skills and knowledge gains they have achieved during the "production process" underpinning their education. Unlike the textbook case in which homogenous labor produces identical outputs, higher education blends the labor of individuals with diverse talents and skills to produce an educated population whose members contribute in a multitude of ways to the economy and to society. Failure to take the quality dimension of productivity into account can lead to suboptimal allocation of resources and incorrect policy conclusions—a point that the literature returns to time and again.

Answering most questions surrounding productivity requires an assessment of the value of inputs and outputs, which requires considering variation and changes in their *quality*.

Measuring changes over time in the quantity, cost, and quality of inputs to and outputs of higher education is essential for monitoring the performance of the sector and to the capacity of colleges and universities to assess and strengthen their long-term economic health—a key objective of the National Association of College and University Business

Officers' Economic Models Project.[1] This review of the relevant literature is intended as a reference for institutions engaged in this complex conversation and attempting to analyze and become more nimble in adjusting the processes underlying the production of higher education.

The remainder of this paper is organized in three sections: Section 2 reviews the conceptual issues that underlie the measurement of productivity and related constructs; section 3 reviews the research attempting to measure aspects of the sector's operations; and section 4 reviews data issues, specifically, gaps in the data infrastructure that constrain the ability of researchers and practitioners to match the concepts with practical measurement.

?

The relevance of a given productivity or performance concept is determined by the question being asked.

2. Economic performance concepts

In higher education, as in other contexts where economic measurement is applied, the relevance of a given productivity or performance concept is determined by the question being asked. Prospective students and their parents may want to know the cost of tuition for a list of schools; but they may also interested in the fields of study offered, the reputation of the faculty, graduation rates and, in some sense, the expected return on their investment—that is, upon graduation, what kind of jobs will be in reach, and what level of earnings might be achievable? Higher education institutions may likewise be concerned about the graduation rates of their students; but they also track many other metrics indicating whether they are maintaining or improving their performance and the extent to which they are competing with peer institutions, both in the United States and abroad—in terms of quality and cost. They are concerned about the quality of students they attract and, ideally, about the performance of graduates they put into the job market, graduate programs, and the world at large. State legislators often focus on per-student costs financed by public coffers, as well as other metrics with which they can assess whether accountability mandates have been met. As a nation, all citizens should be concerned about the extent to which the next generations are being prepared (across a range of disciplines) to drive the engines of economic growth and innovation and to advance societal wellbeing. Meeting these goals depends in large part on the ability of our universities and colleges to maintain the country's research and development capacity, to produce graduates with high levels of human capital, and to foster competition and collaboration with other countries in world markets and in the space of ideas and innovation more broadly. All of these are legitimate foci, and addressing questions embedded in each requires different kinds of information.

1. The project intends to "set a vision for what future economic models might look like and to produce an extensive discussion guide designed to be used by governing boards, presidents, their leadership teams and their stakeholders to engage in the difficult work of structural and cultural change within higher education…[and to offer] a comprehensive tool that provides the foundation for institutions to engage in complex conversations about higher education economic models that are financially sustainable, efficient, effective, and meet the needs of students, employers, and society."

2.1. Costs and other unidimensional metrics

Costs are an objective and, importantly, a quantifiable measure of production processes; indeed, it is often the most emphasized factor among those who monitor the higher education sector. In policy debates, media coverage, and among college and university administrators, attention tends to focus on the soaring sticker price of college as represented by tuition rates. Tuition is the cost to students, but it does not represent the full cost of producing education services because public and other funding sources are also involved. While students are now paying a higher share than ever, overall costs have actually remained more or less in line with general inflation in recent decades[2].

For many questions of interest, costs represent the most visible and most practical metric for quantifying inputs to the higher education production process. In this respect, education is like most other sectors of the economy. Perhaps nowhere is the discussion of costs more prominent than in health care, which (as discussed later) shares many of the same sectoral characteristics as higher education, making its productivity, prices, output, and inputs difficult to measure. A major policy question in health care, as in education, is "Are the returns to investments in them—in terms of quality adjusted life years for the former, and in terms of producing educated citizens and valuable research output for the latter—sustainable and competitive with other high-income countries?" In the case of heath care, the simple (if not always fully correct) answer is frequently that the United States pays more per capita and often gets less in terms of increasing the quality and quantity of life than other industrialized countries (OECD, 2014).[3] Likewise, for higher education, as shown by the OECD data represented in Figure 1, the United States spends more per student than other industrialized countries.

One reason national statistical offices monitor expenditures on higher education spending—beyond the fact that it is one of the few statistics for the sector, along with attainment, than can be harmonized across countries—is that research has clearly established the link between accumulation of human capital (attained through education, formal and informal) and economic progress (Lucas, 1988 and Mankiw, Romer and Weil, 1992). As such, education is a key component of long-run growth. Barro and Lee (2013, p. 186) investigate how output relates to the stock of human capital, finding that "an abundance of well-educated people goes along with a high level of labor productivity…[and] the level and distribution of educational attainment also have an impact on social outcomes, such as child mortality, fertility, education of children, and income distribution." A number of studies (e.g., Romer, 1990, Barro, 1991, and Mankiw, Romer and Weil, 1992) attempt to quantify the link between educational attainment and economic and social outcomes across countries, and some have used school enrollment ratios to answer these kinds of macroeconomic growth questions.

Research has clearly established the link between accumulation of human capital and economic progress.

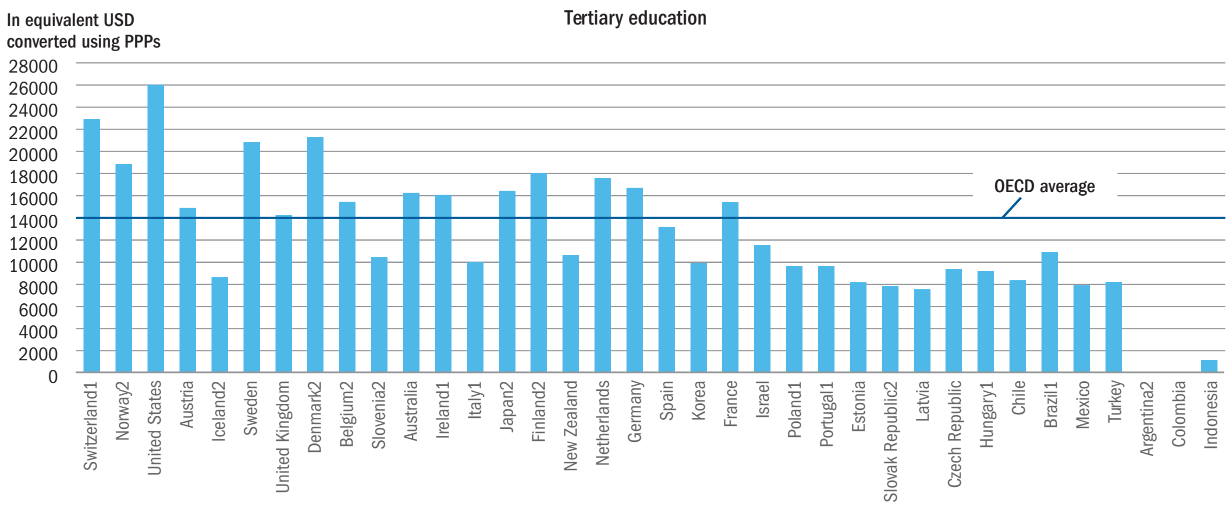
Education is a key component of long-run growth.

2. Garrett and Poole (2006) report that, between 1991 and 2003, tuition increases, adjusted for inflation, averaged 3.4 percent per year at public institutions and 2.8 percent at private institutions, thus outpacing the average annual rate of inflation (which was about 2.5 percent) of other goods and services over the period. Tuition increased faster than overall costs, however, as a higher percentage of institutions' operating costs have been covered by tuition revenues in recent years.

3. The United States spends 16.9% of GDP on health expenditure, more than 7.5 percentage points higher than the OECD country average. And yet the nation's life expectancy at birth is lower than average among the OECD countries. http://www.oecd.org/unitedstates/Briefing-Note-UNITED-STATES-2014.pdf [April, 2016]

Figure 1: Annual expenditure per student for tertiary education (2011)

(in equivalent USD converted using purchasing power parity adjustments so that currency exchange rates may be equalized and compared across countries; based on number of full-time equivalent students)



Source: OECD (2015) Table B1.1a.

Notes: Public institutions only (for Columbia, in tertiary education only; for Italy, except in tertiary education). Countries are ranked in descending order of expenditure on educational institutions per student in primary education.

All too often costs are analyzed in isolation, without being related to the output and outcomes that are being generated.

Cost accounting is one operationally essential way to quantify levels and changes in inputs. But, as alluded to above, costs are only half of the picture in a cost-benefit calculus, and more often than not the easy half computationally. As a result, all too often costs are analyzed in isolation, without being related to the output and outcomes that are being generated.[4] It is possible that countries characterized by higher spending are also those generating commensurately greater benefits. Barro and Lee argue that such aggregate statistics often fail to adequately measure “the aggregate stock of human capital available contemporaneously as an input to production.” Just as it may be defensible to spend more on medical care if it results in greatly improved health outcomes (Cutler, 2004), spending

4. Again, there are analogies with the medical care sector. If costs are generated by inefficiencies or increasing profits of the health care providers, the situation will be viewed very differently compared to that generated by the development of improved treatments and care that results in better quality and longer lives.

more on education may be warranted if it improves learning, human capital and, in turn, the economic and social performance of a country and its population. But, as detailed throughout this review, measuring the outputs and outcomes in such a way that this question can be answered is challenging. The concept of productivity, discussed in the next section, offers a framework for moving forward.

Unit costs

Estimates of the cost of producing an academic credit or degree are relevant for answering questions of interest to a broad range of stakeholders. Unit costing fits into the organizing framework for productivity since it requires documenting all inputs involved in production and then aggregating their costs. It is possible to estimate costs per credit or degree—which, as noted earlier, is very different from the price (tuition) paid for it by students—for academic programs, departments, or institutions, and monitor them period over period such as an academic year or term. The National Governors Association and some state higher education systems have investigated such performance metrics as "credentials and degrees awarded per $100,000 state, local, tuition, and fee revenues—weighted by STEM and health (for example, public research universities in Virginia produce 1.98 degrees per $100,000 received in state, local, tuition, and fee revenues)."[5]

Desrochers (2011) estimated the average cost of producing a bachelor's degree at public four-year institutions to be in the $45,000 to $60,000 range. Brinkman (1985) estimated instructional costs per student credit hour, focusing on differences by field and by course level. Johnson (2009), as part of the Delta Cost Project, constructed various measures of degree production costs for a range of academic disciplines by combining institution-level data on educational expenditures and degrees (collected by the federal Integrated Postsecondary Education Data System [IPEDS]) with state-level data on discipline-level credit-hour costs from Florida, Illinois, and Ohio. The fact that engineering is typically the most expensive major (nearly $100k in fully-attributed costs at four-year public institutions in 2009), but also one that yields very well paid graduates (with starting salaries averaging over $60k for chemical, computer, and electrical engineering), highlights our earlier point that high cost cannot be equated with low productivity or systemic inefficiency.

High cost cannot be equated with low productivity or systemic inefficiency.

5. National Center for Higher Education Management Systems presentation to the Virginia Higher Education Advisory Committee, July 21, 2011.

Notwithstanding the clear need for degree- and credit-based cost metrics, their range of purpose is characterized by essential limitations, several of which were articulated in NAS (2012, pp. 142-143):

> First, aggregating total costs and credits rather than summing from the lowest unit of activity (for example, individual classes) may obscure important differences across micro-environments and may lead to false conclusions because of the disproportionate impact of a few outliers. Second, costs do not necessarily reflect the underlying relationship between inputs and outputs because similar inputs may be priced differently. For example, if one department or institution is staffed largely by tenured faculty with relatively high salaries and another staffed by low-cost adjunct faculty, differences in cost per credit between them may be considerable even though the same numbers of teaching hours are involved. Similar differences encumber cost comparisons across disciplines because of typically high salaries in some (e.g., business) and low salaries in others (e.g., English)…Finally, the cost calculation itself is subject to the joint-use problem because the same faculty member may be doing more than one thing with his or her time.

If the limitations of measuring aggregate total costs are ignored, valuations will become distorted and the policy incentive will always be to substitute low cost inputs for higher cost ones.

The NAS report goes on to caution that, if these limitations are ignored, valuations will become distorted and the policy incentive will always be to substitute low cost inputs for higher cost ones. The report's authors also suggests that statistics on costs per degree and per credit are more appropriate for tracking trends at the national level, where the effects of student and institutional heterogeneity (discussed later in this review) are diluted because they involve averages across large numbers.

If the goal is to know whether increased spending is resulting in commensurate returns, the quantity *and quality* of the inputs and outputs must be reliably tracked.

The problem, then, is not that cost per degree (or graduation rates or retention metrics, discussed later) statistics are calculated and published—they are crucial pieces of information. The problem is when such statistics are presented as though they were comprehensive measures of efficiency or overall productivity. When attention is myopically focused on one dimension of performance, the risk is heightened that goals based on that dimension will be pursued at the expense of others, most worryingly quality.[6] As concluded by the NAS panel (2012, p. 12), "If the aim is to know whether increased spending is resulting in commensurate returns, the quantity *and quality* of the sector's inputs and outputs must be reliably tracked, which, for the latter, requires developing assessment tools for quantifying the outcomes of higher education."

6. Similar tendencies to focus on the easily quantifiable hamper discussions of medical care: the increase in costs is known; however, the value gained from these expenditures, in terms of health benefits to the population, frequently is not.

Graduation rates

Due to their conceptual simplicity and intuitively direct association with the process of higher education, graduation rates are another popular metric for assessing trends related to instructional output. Graduation rates have analogues in other industries: the pace at which cars roll off the assembly line; the rate at which bank loans are repaid; the success of patients completing treatments that improve their health; etc. Likewise, university administrators and funders need to know whether their final product (rather, along with research, health care, entertainment and sport events, etc., one of their many final products) is being produced at the expected rate and on schedule. Statistics produced using the Graduation Rate Survey (GRS), conducted by the National Center for Education Statistics, are staples of accountability reporting in higher education. Graduation rates for four-year institutions are typically computed as the percentage of a starting fall term cohort of first-time full-time students who have completed a bachelor's degree within six years of college entry. The rate for two-year institutions is often specified as a three-year window for completion of an associate's degree. Other graduation rate metrics have been devised that allow the time frame to vary—for example, to account for the presence of part-time students and transfer students in the cohort—or that examine different degree designations. Dropout rates, roughly the inverse (although they have not been defined consistently), are typically calculated by tracking a cohort to observe the percentage of its individuals that is no longer enrolled for at least one credit one year later (NAS, 2012, p. 137).

Although widely used, cohort-based graduation and dropout rates are subject to many limitations. The NAS report (2012, p. 138) reviews these:

> The GRS restricts the denominator to first-time, full-time students, which may represent only a small fraction of beginning students at institutions that enroll large numbers of part-time students and beginning transfers. Including these students in the cohort allows for more completeness, but causes further problems because part-time students have differing credit loads and transfer students bring in wide-ranging numbers of previously-earned credits. This renders fair comparisons difficult because, unlike the first-time full-time population, not all subpopulations are starting from the same baseline. Graduation data, such as that produced by IPEDS, thus penalize certain types of institutions since they do not account for differences in entering students' characteristics or resources available to the college. Graduation rates also reflect admission standards, the academic strength of the enrolled students, and the resources institutions devote to instruction, to remediation, and to retention. Because of this heterogeneity in student types and institutional missions, any increase in the production of graduates (either through increased graduation rates or expanded enrollment) will likely happen at lower ranking institutions; highly selective schools are not going to change and are already operating at capacity.

Along with these limitations, the public policy value of graduation rate measures is compromised because they do not mean the same thing across different types of institutions. As calculated by IPEDS, community colleges have very low graduation rates compared with most four-year institutions. Two-year institutions often provide education to working students who may indeed be less likely to graduate, but they may also take longer to graduate because they are attending on a part-time basis. Perhaps most importantly, many move on to different institutions prior to graduation, such that a transfer (e.g., to a four-year college) should be considered a successful outcome. Even within category, e.g., four-year research universities, or two-year colleges, the range of outcomes in terms of timelines to graduation varies tremendously based on mission. As described in NAS (2012, p. 138), IPEDS graduation rates correctly represent the percentage of full-time, first-time, degree-seeking students who began their post-secondary education at a public college and completed a certificate or degree within 150 percent of the "normal" time at the same institution.[7] They tell us much less about students with "non-traditional" educational plans and trajectories. At the institutional level, open admissions—the key characteristic of the broadest-access schools—are associated with lower graduation rates due to their mission to serve students with educational plans that are more likely to deviate from the four-year convention. These subtleties are generally recognized by researchers, administrators, and other experts in the higher education community, in practice, however, such figures are often condensed in public statements such as "21.5 percent of community college students graduate," which many leaders have come to believe is the entire story.

Responsible use of graduation rate statistics requires additional information on students' background characteristics.

Graduation rate statistics can be useful to students but, as an accountability measure, are fraught with problems of interpretation. Self-selection—specifically, individuals likely to succeed as students have a different set of options than do individuals with low chances of success—means that graduation rates are often poor predictors of how a given individual would perform at one institution relative to others. Also, as a means for monitoring institutional performance, over reliance on graduation rates creates incentives that are not always consistent with mission. If completions were the sole metric by which institutions were graded, incentives would be created to reject applicants with certain profiles—e.g., those who must work or who come from disadvantaged educational backgrounds. If, on one hand, "institutions make failure more difficult by implementing systems of support to help struggling students improve," this is a desired outcome of an accountability system. On the other hand, if they instead "act in ways that dilute a curriculum, or select students who are likely to help improve the institution's ranking," this could be a counterproductive consequence of the system (NAS, 2012, p. 36). Thus, responsible use of such metrics requires additional information on students' background characteristics that are good predictors of success; unfortunately, data on those kinds of background factors are not always available on a large scale to users of graduation rate statistics.

7. https://nces.ed.gov/programs/digest/d09/tables/dt09_331.asp

Graduation rates can be useful for monitoring institutions over time within a specific category—e.g., four-year research universities—at a highly aggregate level. The capacity to answer questions such as "Is the nation graduating a higher or lower percentage of enrollees than it used to?" or "Why is one group in the population succeeding in a certain field of study while another is not?" has value. However, the metric is almost never sufficient on its own—contextual information is needed about students and institutions if graduation rate statistics are to be interpreted correctly. Many institutions have taken steps to provide additional contextual data so that graduation rate information can be interpreted more objectively. The Minnesota state college system, for example, maintains an "accountability dashboard" for each of its campuses designed "as a tool for the Board of Trustees, institutions, policymakers and other visitors…to improve our services to students and to the citizens of Minnesota." It includes student persistence and completion rates, but also indicators such as the pass rates of graduates taking professional licensing exams that provide meaningful measures of quality.[8] The California State University system publishes system-wide information—such as degree-specific estimates of median starting and mid-career salaries, and average student loan debt—to provide a sense of "value to students" of the education provided to graduates from each of its campuses.[9]

Completion and enrollment ratios

Completion rates, based on dividing credentials awarded by the total student population, are another enrollment-based measure of efficiency often applied at institutional levels, but which also may be calculated for state systems, or at the national level. This approach is cross-sectional in that it includes all enrollments; in contrast, cohort-based measures essentially follow students who begin their enrollment at a given point in time. No standard definition is in use across institutions, but a typical calculation divides a count of, say, undergraduate degrees for a given academic year by an unduplicated undergraduate headcount for the same period, or by the number or full-time equivalent enrollments (NAS, 2012). As with graduation rates, completion rates have clear limitations. If student composition is changing, then the population of students captured in the numerator (those completing their credentials) and those in the denominator (the student population broadly) will likely display different characteristics—e.g., levels of academic ability or demographic profiles—that affect their chances of graduating. In such an example, the completion statistic will not be an accurate indicator of the chances of success for the full group. The statistic also will be sensitive to changing population size. If enrollment is growing rapidly, the ratio will understate student success "because the degrees conferred in a given period are awarded to students from an earlier period characterized by smaller entering classes. Some approaches try to correct this defect by counting enrollments four to six years earlier, but the lag amount is arbitrary, so this correction is never entirely satisfactory" (NAS, 2012, p. 140).

Completion rates, based on dividing credentials awarded to the total student population, are another measure of efficiency, but have clear limitations.

8. http://www.mnscu.edu/board/accountability/index.html

9. http://www.calstate.edu/value/systemwide/ [April 2016].

Time to degree

A metric closely related to graduation rates is the average length of time required to earn a degree. This statistic can be calculated for specific degrees at specific institutions or at more aggregated levels. It can be forward-looking, applied to an entering cohort to estimate a graduation rate by averaging the elapsed time from beginning of enrollment to the award of the degree. More commonly, it is backward-looking, calculated for a group of students who were awarded a degree in a given term and then averaging how much time has elapsed since each student's first term of enrollment. The exact specification will dictate whether or not to adjust for things such as students taking a semester or a year off.

As with all metrics, informed interpretation is required to account for factors that influence time-to-degree estimates. For example, it should not count against an institution if it disproportionately grants degrees (such as engineering) involving unusually complex curricula or that require specialized accreditation; likewise, allowances must be made for schools that predominantly serve part-time students, working adults, or underprepared student populations. Uninformed comparisons will result in these institutions appearing less efficient in terms of degree production (i.e., exhibiting longer time values), yet they may be functioning reasonably well given their mission.

Informed interpretation is required to account for factors that influence time-to-degree estimates.

Uninformed comparisons will result in institutions appearing less efficient in terms of degree production, yet they may be functioning reasonably well given their mission.

On the other hand, in some cases, lengthening time-to-degree results could reflect shortcomings in production processes. If an institution has insufficient course offerings and does not deal effectively with students' scheduling complications, a high time-to-degree metric may be indicative of a need for improvement in administrative and operational effectiveness. Bowen et al. (2009) argue that time to degree is an important policy concern that relates to the rising costs of higher education. They find that students who take more than the standard length of time to graduate—either due to starts and stops, switches in majors, or to the schedule of course offerings by the institution—often accumulate large numbers of credit hours beyond what is required for graduation. This can raise both the full cost of a degree and also the cost incurred by students.

In addition to a number of important cost issues, time-to-degree statistics also may provide insights into the broader value of college to students, beyond the amount that they would attach to the degree itself. Very few students attempt to earn an undergraduate degree in three years, while many continue to complete coursework beyond the conventional four-year period. This is suggestive of the presence of benefits created through social interactions and engagements with other students, faculty, and the university community at large—all of which contribute to the learning process. And, for many, the college experience includes a consumption component—as with reading a book or going on a vacation, it may be an enjoyable experience. On the other hand, "when students are pushed to a five-year plan, or if the choice of major and other program options is affected because of insufficient course offerings, this is closer to a productivity problem" (NAS, 2012, p. 141).

In summary, time to degree is a useful piece of information—for example, for student or parents attempting to get a sense of time commitment of various programs and universities. However, it is not a measure that can be used to rank or compare all institutions in a way that necessarily reflects quality or efficiency. As with all the measures summarized here, adjustment factors reflecting student composition must be considered alongside time-to-degree statistics if the goal is to make cross-institution or cross-system comparisons. Even then, it typically makes sense to compare only across like-institutions (for example, community colleges, four-year research institutions, liberal arts colleges, etc.) and to also adjust for factors such as enrollment status and field of degree. Time-to-degree statistics are most useful for tracking the performance of a given institution or department longitudinally, where student characteristics and institutional mission are likely (though not certain) to be fairly stable over time.

Time-to-degree statistics are most useful for tracking performance of a given institution or department longitudinally, where student characteristics and institutional mission are likely to be fairly stable over time.

2.2. Productivity

The limitations of cost and other unidimensional metrics create demands for more comprehensive measures on which to gauge the performance of a firm, a sector, or the aggregate economy. This is where the productivity framework enters the picture. As identified earlier, productivity is commonly defined as a ratio between output (the goods or services produced) and inputs (such as labor and capital). The input-output relationships may be expressed in terms of production units—e.g., hours of labor and numbers of degrees conferred; or it may be expressed in terms of dollars—e.g., the cost of inputs required to graduate a student and the value to students and society of the degree. Measures of productivity may be calculated at the micro level and at the macro level, from firms or institutions all the way to national aggregates. The broadest, and one of the most widely used, measures of productivity is Gross Domestic Product (GDP) per hour worked. In higher education, true productivity measurement requires assessing quality-adjusted output (credit hours, degrees, etc.) per quality-adjusted unit of input (faculty, staff, laboratory equipment, etc.).

True productivity measurement requires assessing quality-adjusted output per quality-adjusted unit of input.

The NAS report (Chapter 2) goes into great detail defining the "production function" for higher education. The focal points of the discussion are on labor productivity (the ratio of output per labor-hour), and multi-factor productivity, which relates a measure of output to a bundle of inputs (e.g., labor, capital, and purchased materials). Productivity increases when more output is produced with the same inputs or when the same output is produced with fewer inputs. Often, the question of interest—e.g., are individual students, and society broadly, getting more (or less) for their investment in higher education than they were 20, 30, 40 years ago?—requires estimating how this relationship is changing through time.

The comparatively simple metrics described in the previous section are regularly used as proxies for the productivity or performance measurement concept of interest, which is often unattainable. Partial measures should not be equated with productivity. If a university is graduating more of its students, or if it has held tuition or cost per degree in check, it does

Some studies claiming to measure productivity actually measure efficiency or unit cost.

Blurring cost and productivity has led to a disproportionate focus on reducing the cost of credit hours or credentials.

Attention to output quality requires a

focus on students.

not necessarily follow that it has increased productivity. Belfield (2012) found that some studies claiming to measure productivity at four-year institutions were in reality measuring efficiency or unit cost, such as the cost of producing a graduate. And, indeed, cost savings may arise as a result of a reduction in input prices or from more efficient use of inputs in the production process such that the same physical quantity of outputs are being produced at a lower total cost; however, cost savings also may be attainable by reducing the quantity or quality of output.

The NAS (2012) report argues that a blurring of cost and productivity discussions has led to a disproportionate focus on reducing the cost of credit hours or credentials, which invites the usual solutions: substituting graduate students and other low-cost teachers for expensive, seasoned faculty; increasing class sizes; and eliminating departments that serve small numbers of students. Likewise, Archibald and Feldman (2011:40) write:

> An institution can increase class size to raise measured output (students taught per faculty per year) or it can use an increasing number of less expensive adjunct teachers to deliver the service, but these examples of productivity gain are likely to be perceived as decreases in quality, both in the quality rankings and in the minds of the students.

Thus, while costs per unit of output is an essential performance measure, if all stakeholders cared about was cutting costs and quantities, the job of assessing performance would be done—we already know how to do these things. In the economic textbook case in which a firm manufactures homogenous widgets using homogenous labor, costs fully reflect efficiency and productivity. However, when output is multidimensional and widely varying in quality, much more information is needed in order to assess economic performance. As described by Massy et al. (2012), quality—which is the most difficult aspect of productivity measurement on which to collect data and to measure—is “the elephant in the room.” If quantifying inputs and outputs, or costs, was all that was needed, we would not be asking the same questions today that were being asked 100 years ago in the Carnegie Foundation Report (Cooke, 1910) discussed in the next section.

Attention to output quality requires a focus on students. Even if the objective of a particular exercise is to compare performance across institutions (say, for a Carnegie classification peer grouping), estimating the net impact of attending college—whether expressed in terms of learning gains or by the expected marginal benefit on graduates’ lifetime earnings—inherently requires that the student be the base unit of measurement. This is a central point in the paper by Nate Johnson—writing for the TIAA Institute Higher Education Series: Understanding Academic Productivity—in which he argues that the way productivity is typically measured within postsecondary institutions can be helpful for institutional management purposes, but can yield misleading results when generalized to a metropolitan area, state or nation. Much of the key information needed for developing policies for the optimal allocation of funding—for example, financial aid policy for public institutions—requires student level data. For accountability or performance measurement purposes, this data should be aggregateable to the level of governance (local, state or national) that is relevant to the question at hand. Policy makers, Johnson notes, also should factor student inputs as well as institutional inputs in their analytical framework and should make distinctions primarily among categories of students rather than categories of institutions when they disaggregate data (Johnson, 2016).

If higher education is to become more productive, institutions must generate output of a greater value with given resources, or produce the same quality output at lower cost. Technology is often the route through which various sectors of the economy have increased productivity, whether measured in terms of output per hour of labor or multifactor productivity. While there can be little doubt that use of information technology (computing power, internet resources, etc.) makes it possible for instructors, researchers, and students to accomplish more with their time, the evidence on the potential of higher education to benefit from new models of production, such as online courses, is far from conclusive. Harris and Goldrick-Rab (2011) argue that "researchers and institutions themselves have rarely paid much attention to whether policies and practices are cost-effective. How would you know whether you're spending money effectively if you've never even asked the question?" They conclude that colleges "can conceivably become more productive by leveraging technology, reallocating resources, and searching for cost-effective policies that promote student success." Indeed, many industries such as retail trade and banking and finance that formerly were believed to be stagnant have been able to improve productivity dramatically (Triplett and Bosworth, 2006).

Over the long run, for the economy as a whole, productivity growth—often driven by changing technology (most recently in information, communications, and computing)—has been the primary source of increases in living standards (Bosworth, p. 61). On the other hand, cutting costs is rarely cited as the key to boosting economic growth, incomes, and increasing standards of living. When more can be produced with the same amount of resources, per capita income rises. Likewise, for specific sectors, especially those that rely to some degree on stagnating or declining public funding resources, the most feasible route to improved cost efficiency or increased output is often through productivity growth. The NAS report (p. 19) acknowledged this point: "In higher education, productivity improvement is seen as the most promising strategy for containing costs in the continuing effort to keep college education as affordable as possible." The report goes on to predict that, "without technology-driven and other production process improvements in the delivery of service, either the price of a college degree will be beyond the reach of a growing proportion of potential students or the quality of education will erode under pressures to reduce costs."

Productivity improvement is seen as the most promising strategy for containing higher education costs.

An ideal productivity measure—one that accurately calculates the quantity and quality of all inputs and outputs—would answer many of the key policy and operational questions that now confront higher education, such as whether the return on the nation's or a state system's use of public funds is increasing (or decreasing), and how that return compares with other countries or state systems.[10] However, even if such a perfect measure existed, it could not inform every policy decision. For example, productivity measures do not address allocative efficiency—the optimal amount of education that should be produced. They do not tell us if society spends too much on defense or other options and too little on education, or vice versa; these questions require conversations of a political nature. Nor do productivity measures shed light on the question of *who* should pay the cost of education. The share of the expenditures carried by students and parents consuming the service on one hand and by public finances on the other, reflects, at least indirectly, a view of the purpose and desired outcomes of the enterprise. Investment by individuals in education is justified by the private return (which, depending on the consumer, might be earnings, satisfaction from obtaining knowledge, building networks of friends and contacts, or some combination of all of these) that accrues to those earning credit hours and degrees. The rationale behind public investment is founded on a recognition that education generates positive externalities, or benefits to society, beyond those that accrue directly to the consumer (in this case, students). To the extent that an educated citizen is more likely to innovate, be civically engaged, and pay taxes, and less likely to commit crimes or be supported by the state, everyone should contribute to this preferred social and economic outcome.

An educated citizen is more likely to innovate, be civically engaged, and pay taxes, and less likely to commit crimes or be supported by the state.

Everyone should contribute to this preferred social and economic outcome.

2.3. Efficiency and effectiveness

The concept of efficiency is closely related to that of productivity, and its measurement has long been pursued, even in the context of higher education. Going back more than a century, at a time when public institutions already were being asked to justify their "rate of return" to the public, the publication *Academic and Industrial Efficiency: A Report to the Carnegie Foundation for the Advancement* (Cooke, 1910) took as its charge to "measure the efficiency and productivity of educational institutions in a manner similar to that of industrial factories" (Barrow, 1990:67). Authors of the report developed a time-use accounting formula to estimate the costs and outputs of higher education for teaching and research; the unit of measure on which the productivity measure was to be based in that report was the student hour, defined as "one hour of lectures, of lab work, or recitation room work, for a single pupil" (ibid, p. 70). The initiative was motivated by a perceived managerial need for metrics on relative faculty workloads and on the cost of instruction per student hour, which could be used for the purpose of comparing the rate of educational efficiency for individual professors, fields, departments, and universities (Shedd, 2003).

10. Measurement of inputs, outputs, productivity and price has come closer in our economic statistic to the ideal in some sectors that are less amorphous than higher education. At the Bureau of Economic Analysis and the Bureau of Labor Statistics, today's cars and computers are not directly compared to those produced ten or twenty years ago: When measuring price and productivity changes in these sectors, techniques are now employed that account for changes in input and output characteristics. For example, when airbags where introduced, the quality improvement component of the price change was not attributed to price inflation, but to the fact that the car was worth more with an added life-saving feature.

Practical considerations regarding the operation of a college or university dictate an efficiency mindset—that is, maximizing the number of students that can be educated with a fixed amount of faculty, staff, and facilities—at least in the short run. Efficiency is a key concept for addressing questions about the quantity and quality of educational programs being offered by an institution given budget constraints. Focusing on outputs, Farrell (1957) defined efficiency as the ratio of a firm's observed output to the maximum output that could be achieved given its input levels. This measure of efficiency, therefore, requires estimating a production function—the engineering relationship between inputs and outputs—in order to compare the potential maximum production point with the observed production point. Similarly, Kokkelenberg et al. (2008, p. 2) described three aspects of efficiency: "allocative efficiency which means the use of inputs in the correct proportions reflecting their marginal costs; scale efficiency which considers the optimal size of the establishment to minimize long-run costs; and technical efficiency which means that, given the establishment size and the proper mix of inputs, the maximal output for given inputs under the current technology is achieved." In the context of higher education, allocative efficiency dictates, for example, whether an institution has the right number of classrooms and instructors given the number of students enrolled. Scale efficiency has to do with how per-unit (average) costs change along with the size of the institution. Technical efficiency is close to the version defined by Farrell: what is the maximum amount of education (e.g., number of credit hours or degrees, quality of credit hours or degrees, or a combination of the two) that can be delivered given current levels of labor, capital, and other inputs of production. In contrast, effectiveness, as defined by Garrett and Poole (2006) might be reflected in, for example, the employment rates or quality of job placements of recent graduates.

Garrett and Poole (2006) describe how productivity and efficiency, along with effectiveness, interrelate:

> A more thorough definition of productivity recognizes that productivity can be divided into two parts: efficiency and effectiveness. Efficiency refers to the level and quality of service that can be obtained given an organization's fixed resources. Thus, an organization is considered more efficient if it can increase the level or quality of service without increasing the amount of inputs used. Effectiveness, on the other hand, refers to how well an organization meets the demands of its customers. The customers in higher education are students, parents, employers and state legislatures. Customer demands may include such outcomes as a specialization of knowledge in a specific area, career assistance and job placement and, probably most important, the graduation of well-educated and productive students.

Productivity can be divided into two parts:

Efficiency | **Effectiveness**

The NAS report (2012, p. 32) describes efficiency is a similar manner:

> Efficiency is improved when cheaper inputs are substituted for more expensive ones without damaging quality proportionately. When institutions substitute adjunct instructors for tenure-track faculty, costs are certainly reduced. Whether this move toward lower-priced inputs has a proportionately negative impact on output quantity and quality (e.g., numbers of degrees and amount learned) is not yet fully known, and surely varies from situation to situation (e.g., introduction and survey classes versus advanced seminars).

On the issue of the labor-input mix, Ehrenberg's (2012) review of the evidence leads him to conclude that, in a wide variety of circumstances, the substitution of adjuncts and full-time non-tenure-track faculty for tenure-track faculty has resulted in a decline in persistence and graduation rates.

The efficiency measurement concept becomes stretched when the functional composition of the labor input is allowed to shift over time, because it is difficult to detect quality changes that accompany it. At a level of administrative organization higher than tenure/non-tenure input mix choice, Garrett and Poole (2006) find that instructional expenditures at public universities, expressed as a percent of total expenditures, have decreased from 39 percent in 1977 to 34 percent in 2001, while administration expenditures increased from 30 percent of instructional expenditures in 1976 to 50 percent in 2001. Further, they found that, "while inflation-adjusted instructional expenditures per student increased by 17 percent between 1990 and 2001, administrative expenditures per student increased by 54 percent over the same period." On the surface, this sounds like an ominous trend, but without a parallel measure of the impact of this resource shift on the quality (or even the quantity) of degrees and research or other outputs produced by these inputs, it is not possible to convincingly make the case that this shifting of personnel has or has not been detrimental. Webber and Ehrenberg (2010) examined one facet of the administrative expenditure, outcome quality link. They found that, for institutions with high numbers of low-income students (based on Pell Grant eligibility) and with lower entrance test scores, higher administrative expenses on student support services were positively associated with graduation and first-year persistence rates. The authors recommended further research to determine which among the many subcategories of student service expenditures contribute most to improved outcomes.

When the focus is on output and input volumes alone, **it becomes difficult to distinguish efficiency gains from quality changes.**

That the quality of education should be fixed in an analysis of efficiency is easier said than done. When the focus is on output and input volumes alone, it becomes difficult to distinguish efficiency gains from quality changes. For example, Babcock and Marks (2011) reported that college students were, at the time of their study, studying fewer hours than they did in previous decades.[11] Assuming studying is an input to learning, does this mean that students have become more productive because they are passing classes with fewer hours work, or does it mean they are now studying less and learning less? Arum and Roksa (2010) argue that college students are learning less. But, without robust time series data on test results and other outcomes measures to verify student learning, the question remains unanswered (NAS, 2012).

A particularly relevant and timely example of resource maximization is the use of technology to optimize classroom structure and course design. Throughout the economy, productivity improvements have frequently been identified with technological change, but they also may be associated with iterations toward best practice as a means to reduce inefficiencies. Several organizations operating within the higher education space have taken this approach recently. The National Center for Academic Transformation (NCAT), for example, is a nonprofit organization that describes itself as "dedicated to the effective use of information technology to improve student learning outcomes and reduce the cost of higher education."[12] NCAT works directly with institutions to analyze their instructional models and identify ways to efficiently leverage technology to improve efficiency in production of higher education through course redesign. Reflecting this mindset, the prescription for improving productivity advanced in Garrett and Poole (2006) involves better alignment of faculty and student needs: "At many universities, as student demand for certain majors or classes ebbs and flows over time, there is little change in the number of faculty in each department. A failure to match teaching capacity with student demand is completely opposite the private sector, where changes in business conditions directly influence staffing levels." Thus, beyond technology and course design, one tool for increasing efficiency and cost effectiveness is in changing the way colleges organize programs and support systems along the student's pathway toward graduation.

3. Measurement

The discussion of measurement concepts above provides a sense of the kinds of questions that can be answered using various cost and performance statistics. Operationalizing the concepts into useful, comprehensive measures of productivity requires data, of course, but also an understanding of the specific characteristics of higher education, some of which are shared with other service sector industries. In this section, the unique attributes of higher education are examined, along with how these attributes factor into measurement of the sector's activities at different levels of aggregation, from individuals and departments up to institutions, systems and nationally. Research and data platforms that attempt to put the concepts into practice to measure inputs and outputs in the education production process also are reviewed.

11. Babcock and Marks (2011) use data from the American Time Use Survey, a module of the Current Population Survey conducted by the Bureau of Labor Statistics as their source of information on student input hours. Time use studies also indicate that students engage in patterns of homework and other time use behaviors at different levels of intensity depending, for example, on their majors or enrollment status.

12. http://www.thencat.org/

3.1. Characteristics of the service sector of the economy

Service-producing industries have emerged as the dominant engines of U.S. job and productivity growth in recent decades. This has not always been the case, however, and this development was slow to be recognized. For many years, a conventional view of the role of service industries in productivity growth of an economy was that portrayed in the landmark book, *Performing Arts: The Economic Dilemma*, by Baumol and Bowen (1966). In it, the authors put forth an example of why it is difficult to improve productivity in service sector industries, which are often labor intensive. The example, as retold by (Bosworth, p. 62) is that "in 1787, when Mozart composed his String Quintet in G Minor, it took five musicians to perform it. Today, over 200 years later, it still takes five people and, unless they play really fast, it takes as long to perform the piece as it did centuries ago. In contrast, the resource cost of producing most goods has steadily declined." Archibald and Feldman (2011) point out that education shares some of the characteristics with the string quartet: the traditional production processes for colleges and universities rely on human interaction and nearly fixed amounts of time inputs from faculty and students. In higher education, it also may be true that workers produce about as much per hour as they did 10, 20, or 50 years ago if an average college instructor cannot grade papers or give lectures any faster now than then.

In industries where productivity gains do not take hold, **costs keep going up.**

This cost problem for labor-intensive industries, as many in the service sector are, arises in response to the impact of increasing productivity on overall wage trends in the economy. Classical labor market theory predicts that wages will be driven up in sectors benefiting from technological advances that increase labor productivity. These sectors, however, may not experience increased per-unit production costs because the advances allow them to make more with less (for example, it takes roughly half as many labor hours to build a car now compared with 30 years ago). Increased production of output generated by each unit of input offsets the increased wages paid to labor. Meanwhile, sectors that cannot take advantage of new productivity-improving technologies will still face a higher wage bill because they must compete for workers in the marketplace for labor; yet, they will not benefit from the efficiency gains. Additionally, if the industries that have become more efficient through labor-saving technologies or other productivity-enhancing means are able to employ a lower ratio of labor to capital in production, the impact on labor intensive industries will be proportionately even greater. The result is that, in industries where productivity gains do not take hold, costs (and the prices of their goods and services) keep going up.

Recent history, however, suggests that some seemingly entrenched production processes have overcome Baumol's *cost disease*, as it came to be known. The financial sector is a good example, where the move to electronic trading for many financial instruments has taken hold in what was once a process characterized by direct human-to-human interaction. In a way, even the Mozart bottleneck has been resolved: "In the music industry, while not everyone can afford tickets to hear a quintet live, for about $15 they can buy a recording of that quintet that is likely to be of higher quality than what they would hear from most of the seats in the music hall" (Bosworth, 2006, p. 63). String quartets have improved "productivity" dramatically through the capability to simulcast performances and to record music, allowing money to be earned while the musicians sleep (Massy, 2010:39). Even in aggregate economic statistics—where observed cost-disease may be as much an artifact of measurement problems (because outputs are difficult to define and measure) as it is of real productivity deficiencies—Bosworth finds that things have changed. Improved industry-level input/output accounting data indicate that, contrary to Baumol's example, service sector industries have actually led growth in labor and multifactor productivity in recent decades through use of IT and other high tech capital and through process innovations.

Service sector industries have actually led growth in labor and multifactor productivity in recent decades.

That production processes are in fact changing in higher education should not come as a huge surprise; intuitively, education is an activity that should benefit a great deal from advances in information—since that is what learning is all about—and other new technologies. At a granular level, the work by NCAT referenced above has begun showing measurably improved learning outcomes and, in some cases, reduced costs generated through effective use of information technology and revised instructional models. These are early indicators that higher education will ultimately follow the path of other service sector industries which, bucking the gloomy prognosis implied by Baumol, have emerged as the most dynamic and innovative sectors of the U.S. economy. This was the verdict of Triplett and Bosworth (2002), who proclaimed that "Baumol's disease has been cured" and that the characterization of services as a drag on aggregate growth is no longer valid. These overall trends in the service sector suggest that there is optimism that the education sector may be able to stabilize cost growth and the rapid inflation of tuition.

The education sector may be able to stabilize cost growth and the rapid inflation of tuition through efficient use of information technology.

As the service sectors—and particularly industries such as health care, education, and finance—have grown as a share of the U.S. economy, so too has the need to improve measurement of them. Some part of the growing awareness of the leading role of services today in creating jobs and raising productivity has arisen as a result of improvements that have been made in data collection and measurement by statistical agencies, and by businesses and institutions. As described above, productivity measurement is comparatively easy when a physical item is being produced, since one need only count the amount of the output being produced by each unit of input. Economic models may closely resemble real-world production processes in simple manufacturing cases in which the good being produced is easily identifiable and easily quantifiable. Measurement is further eased when the good or service is transacted in markets such that price can be equated with value, reflecting both quantity and quality dimensions. Services such as education and health care do not exhibit these characteristics and, as a result, are among the most difficult sectors to measure.

As observed in NAS (2012, p. 22), "it is possible to count and assign value to goods such as cars and carrots because they are tangible and sold in markets; it is harder to tabulate abstractions like knowledge and health because they are neither tangible nor sold in markets." The difficulty of measuring productivity in service industries has been well documented. Sherwood (1994) and Dean and Kunze (1992) both point out that a key stumbling block has to do with the ability to identify and define the basic unit of output. In medical care, is it an hour of doctor's time, a day in the hospital, or neither? Consumers (patients, in this case) do not demand hours in a hospital or blood transfusions—nobody *likes* these things. Rather, they typically hope to purchase a completed, and hopefully successful, treatment; they want to feel better. But data are not always organized to track costs and prices at the relevant transaction level. A hospital may have data only on the cost of the doctor's time or the daily price of a hospital bed; the information may not be organized by the relevant transaction unit, such as a full treatment.[13] Similar problems exist for education: is the unit of output an hour in the classroom, a completed course, a degree? The answer depends on the question being asked, and the ability to collect data on the relevant units of inputs and outputs. To that end, accurate and detailed information is needed on levels of capital investment; employment levels; and, perhaps most importantly, about the range of outputs, their quality, and their value.

3.2. Unique attributes of higher education

The higher education sector is endowed with characteristics that are difficult to quantify using standard microeconomic methods: colleges and universities produce multiple heterogeneous outputs from multiple heterogeneous inputs; they are often nonprofits; market-determined prices are absent; and many of the contributions of higher education to the economy and to society are non-market or public in nature. Here, we discuss each of these aspects of the sector.

Multiple heterogeneous inputs and selection issues

Diversity of inputs and outputs is not unique to higher education, but the extent to which it characterizes its components—the people, activities, processes, and transformations that occur—is unique. On the input side, heterogeneity is found in students of varying abilities and ambitions, teachers and administrators of different skill levels and areas of expertise, and institutional organizations that are driven by a wide range of priorities and missions. Understanding the relationship between the inputs and outputs of higher education is further complicated by selection issues. If students were all of identical quality—or even if they were not but they were distributed randomly across all colleges and universities—it would be easy to assess and compare the average gains in learning and in job preparedness as a means of assessing quality across departments, institutions, or systems. But students with different aptitudes and objectives systematically go to different schools. A similar diversity of skills and interests exists among instructors, the key labor input to learning. This characteristic of the inputs of higher education has been shown to be an important factor in productivity and performance analyses and, as reviewed below, becomes especially clear in findings from value-added models.

13. Progress is being made on price and productivity measurement in health care. The Bureau of Economic Analysis has been working for the past decade on producing disease and treatment-based health care accounts and price indexes (http://www.bea.gov/national/health_care_satellite_account.htm).

Multiple heterogeneous outputs and joint production

Higher education cannot accurately be characterized as a single industry; rather, its countless activities contribute to the output of many industries. Institutions of higher education are multi-product firms, delivering instructional programs that result in educated citizens, research findings, entertainment in the form of cultural and athletic events, housing, medical care, and community services. These offerings contribute to the vibrancy and livability of the college towns in which they reside and to places beyond. In order to measure the production function of higher education, one must disentangle the joint production of these many outputs and identify the contribution of each input used in their production. For market-based enterprises, the production of multiple products raises estimation issues that are manageable, because outputs can be combined on the basis of their relative contributions to revenue shares. But the most obvious output of higher education, educated graduates, are not sold in a market, although their skills and credentials may be brought to the labor market where they are valued. Education creates human capital (knowledge and skills) that is expected to increase economic productivity, which is priced in the labor market; hence college graduates earn more on average than non-graduates.

Higher education cannot accurately be characterized as a single industry.

To measure the production function of higher education, one must disentangle the joint production of many outputs and identify the contribution of each input used in their production.

In the university, an instance of joint production that comes immediately to mind occurs when faculty spend time engaged in both instruction and research. These may be complementary activities (when engagement in research makes the professor more informed, more inspirational, and a better teacher) or substitutes (when the professor spends time on research at the expense of teaching quality). NAS (2012) focuses on the instructional component of higher education productivity—described as all taught programs, regardless of level (e.g., associate, bachelors, masters)—but acknowledges that “it is not simple to isolate one component when joint production of instruction, research, and public service is occurring simultaneously” (p. 14). When missions are budgeted separately, as is sometimes the case with health care and athletics, the calculation is somewhat easier, although “synergies no doubt exist between these activities and instruction programs.” Research, on the other hand, is especially closely linked to the education mission. The value of research and of the universities that house it cannot be fully measured by their instructional contribution alone: “Important interactions exist, both positive and negative, between research activities and the productivity of undergraduate instruction. On the positive side, there is the opportunity for promising undergraduates to work alongside experienced faculty. On the negative side, the possibility exists that the growth of graduate programs detracts from commitments to undergraduate education” (NAS, pp. 22-23).

The link between research and instruction is a perennial topic of interest.

Even measuring research productivity on its own, if one could isolate the hours with and without a connection to instruction, is far from straightforward. Wexler (2015) asks, "Can Data Measure Faculty Productivity?" and then explains why a "Rutgers Professors Say No." Data that can be used to assess research productivity, such as that compiled by the company Academic Analytics, includes variables such as "professors' journal articles, citations, books, research grants, and awards, and compares those numbers with national benchmarks." The problem is that these factors, even if accurately tracked, must be weighted subjectively if the output associated with them is to be quantified or given a valuation. Not all journals have equal credibility and influence, and no one can predict the long-run, or even short-run, impact any given research finding will have. Further, if we bring joint production back into the picture, it is difficult to assess the extent to which research complements instructional effectiveness on a case-by-case basis.

The relationship between research and instruction is a perennial topic of interest, largely due to the perception that, in some institutions, research is prioritized beyond what is warranted in terms of mission, and that there is a negative correlation between expenditures on research and student outcomes. Webber and Ehrenberg (2010), analyzing data assembled by the Delta Cost Project (Wellman, 2008), confirm that in recent decades "instructional spending per full-time equivalent (FTE) student in both public and private four-year colleges and universities in the United States grew at a slower rate than median expenditures per FTE student in many other categories of expenditures (research, public service, academic support, student services, and scholarships and fellowships)." Further, using institution-level data from IPEDS, the authors found evidence that higher levels of expenditures per student on budgeted research expenditures were associated with lower graduation rates. Though the data were inadequate to confirm it, they speculated that this relationship arises "because institutions with higher levels of budgeted research are also institutions in which a greater share of instructional expenditures are devoted to departmental research" (p. 18).

Ramsden and Moses (1992) examined the associations between research and undergraduate teaching in Australian higher education. Using a research index based on publications and other activities and an index of teaching effectiveness based on reported commitment to teaching, the authors concluded that there was "no evidence …to indicate the existence of a simple functional association between high research output and the effectiveness of undergraduate teaching. The practice of allowing research performance to act as surrogate for teaching performance, as often happens when lecturers are appointed, is insupportable; teaching and research need to be separately assessed" (p. 273).

Porter and Toutkoushian (2006) examined the connection between research productivity and student quality and overall institutional reputation. They posited that an "institution's reputation, research output, and average student quality are determined simultaneously… because these outputs are produced jointly." They found that "faculty research productivity is positively related to reputation but negatively related to student quality at research universities, but that reputation and student quality have little impact on research productivity at liberal arts colleges." Their findings suggest that institutions may face a tradeoff between "having a high quality research faculty and a high quality undergraduate student body" (p. 605). A negative relationship between research and learning (if it truly exists) would suggest a need to find ways of continuing research programs without creating negative impacts on instruction and the probability of success for students.

It is important to acknowledge that research creates social and economic value far beyond what can be captured by its effect on graduation rates.

While findings on the joint production relationship between teaching and research raise legitimate policy concerns, it is important to acknowledge that research creates social and economic value far beyond what can be captured by its effect on graduation rates, which is surely a second-order effect. At the margin, even if true, some decrease in graduation rates would be a price worth paying if the research yields medical advances that save lives, technological innovations that lead to long-run economic growth, and contributions to general science that may bear fruit for years to come. Additionally, there are long-run requirements for maintaining and building upon the stock of human capital and teachable knowledge. If research were curtailed, instruction might improve locally and in the short run, but research is what generates the body of knowledge that is taught for generations to come. Therefore, for some questions, the cost of research and the value of its output must be assessed independently of its contribution to the production function for instruction narrowly defined. Dundar and Lewis (1998) examined the relationship between academic research productivity and institutional factors in the biological sciences, engineering, the physical sciences, mathematics, and the social and behavioral sciences in the nation's research universities. They found that institutional attributes and departmental culture and working conditions affect research productivity. For example, their data suggest that larger departments are more likely to reach the critical mass necessary to stimulate collaboration and nurturing of research activity.

The existence of multiple outputs reflects the wide-ranging goals of higher education, strongly suggesting a need for **multidimensional performance measures.**

In summary, the existence of multiple outputs reflects the wide-ranging goals of higher education, which as argued throughout this paper, strongly suggests a need for multidimensional performance measures. As expressed by Cunha and Miller (2014, p. 66), "a full appraisal of institutional performance at the postsecondary level likely requires a set of indicators that proxy for the various dimensions of institutional performance, and this set of indicators is necessarily limited by data availability."

The contribution of higher education to the public good, and in the provision of nonmarket outputs

Outputs and outcomes produced by colleges and universities, other than those that are easily countable (degrees and credit hours) and market oriented (e.g., the sheepskin or credentialing earnings return[14]), have an important place in the productivity conversation. Many of the benefits of education manifest as positive externalities—that is, not all of the value associated with higher education output accrues directly to those who have paid for it or who even those who are directly involved. Some of these spillover benefits are pecuniary and market based in nature, but even more are not.

Outputs and outcomes produced by colleges and universities have an important place in the productivity conversation.

Moretti (2004) estimates positive spillovers of education in labor markets using evidence from longitudinal and repeated cross-sectional data. By comparing the wages of similar individuals who work in cities with different shares of college graduates in the labor force, he shows that a percentage point increase in the supply of college graduates raises high school drop-outs' wages by 1.9%, high school graduates' wages by 1.6%, and college graduates wages by 0.4%.[15] But the positive spillovers of higher education extend far beyond the market. The sector contributes to a citizenry that is healthier and better equipped to contribute to the public good—through civic engagement, contributions to public finance, being less likely to commit crimes, etc.[16] Matthew Lambert explores these and relates themes in his book, *Privatization and the Public Good: Public Universities in the Balance wherein he concludes that both the* individual and societal benefits of higher education are seriously undervalued. Likewise, the *New York Times* columnist Frank Bruni points to a number of the issues described above in an opinion piece, "College's Priceless Value" to explain why higher education is not, and should not, be solely a private business, and why higher education rightfully has been considered as belonging to the public sector. Higher education provides knowledge, and knowledge is a public good. Beyond this, high levels of education are associated with the capacity to make more informed life decisions—about health, parenting, etc.—that have been shown to be linked to self-reported well-being (Oreopoulos and Salvanes, 2011). Walter McMahon's (2009) study found that, of the total benefits of higher education, which he argues are grossly underestimated, more than half of the benefits accrue to society as positive externalities and public goods.[17] Tilak (2008) estimates that the diverse and numerous spillover benefits (monetary and nonmonetary) from higher education to society as a whole are even greater than the already high private benefits.

Higher education provides knowledge, and knowledge is a public good. **High levels of education are associated with the capacity to make more informed life decisions.**

14. Even within the labor market, there are non-pecuniary returns to education, in the form of greater choice in type of jobs for which a person is able to qualify.

15. A key methodological issue that must be dealt with in these kinds of cross-area comparison studies is the presence of unobservable characteristics of individuals and the specified geographic locations (cities, in this case) that may raise wages and be correlated with the explanatory variable (college share in this case). Instrumental variables are often used to establish causal relationships when the selection issues are present.

16. The Association of American Colleges & Universities' VALUE Rubrics (Valid Assessment of Learning in Undergraduate Education) specifies a set of learning outcomes in the areas of "Intellectual and Practical Skills," "Personal and Social Responsibility," and "Integrative and Applied Learning" for assessing civic engagement—and specifically how well institutions prepare "graduates for their public lives as citizens, members of communities, and professionals in society" (https://www.aacu.org/civic-engagement-value-rubric, 2016).

17. The formal definition of a public good is that it is "non rival" and "non excludable" in that the good or service being consumed by one individual does not reduce its availability to others, and no one can in principle be excluded from its consumption. National security and street lighting are textbook examples.

In assessing the contributions of higher education, failure to capture these broader benefits seriously distorts value added and return on investment calculations which, in turn, may lead to misallocation of public funds and sub-optimal decision making by state funders and others. Unlike most profit-maximizing firms, institutions of higher education share a broader objective function that is concerned with nonmarket outcomes. Objectives of the university may include increasing student quality, increasing access and diversity, achieving greater cost-efficiency, making a greater contribution to the needs of the community, and improving basic research. Of course, other firms and organizations contribute to the greater good, but this is effectively the mission of colleges and universities.

If broader benefits are not captured in assessing the contributions of higher education, value-added and return on investment calculations can be distorted.

The NAS report (2012, p. 12) weights in on the non-pecuniary, nonmarket, and public goods aspects of higher education in arguing that standard productivity statistics, while vital, will always be incomplete:

> There are aspects of human and, more narrowly, productive enterprises that create social value but that statistical measures do not and indeed do not presume to capture. From a societal perspective, investment in citizens' work careers is not the only motivation for supporting and participating in higher education. Nonpecuniary components of the sector's output associated with instruction, research, and other public goods are also important. Like a policeman who brings extraordinary passion to protection of fellow-citizens, a technology entrepreneur whose vision ultimately changes the way people live, or an artist who is appreciated long after creating the art, the passion and dynamism of a master teacher who is truly interested in a student who, in turn, is truly interested in learning cannot be richly portrayed in a number. In this context, some very real elements of the value of experiencing life-changing learning cannot be fully quantified within a (still very important) statistical infrastructure.

That the role of the higher education sector in contributing to the production of public goods is being given extensive attention—in the papers noted above, in the NAS report, and elsewhere, is a clear indication of how the conversation about higher education performance has evolved. A forthcoming edition of *Higher Learning Research Communications* is inviting "scholars, administrators, and public policy officials to explore the role that colleges and universities can, should, and must play in addressing a shared public good for the benefit of local communities and for the preservation of our global shared space and future." Papers are expected on a wide range of "nonmarket" topics intended to draw further attention to the impact of education on outcomes in areas as diverse as economic mobility, income inequality, climate change, sustainable growth, and human and social welfare generally. Genevieve Shaker and William Plater—writing for the NACUBO Productivity Series of which this review is part—have written two essays on "The Public Good, Productivity, and Purpose: New Economic Models for Higher Education." The first focuses on goals associated with the use of resources to "achieve both institutional mission and to ensure a civic return on the investment of public funds." A second paper by the authors examines "how *faculty* in particular might be held accountable for individual contributions to the public good work of their employers." Among the takeaways from their research are that contributions to the public good cannot be adequately accounted for in exclusively economic (monetary) terms; metrics to assess the public good necessarily must focus on educational outcomes, and civic learning specifically, at the institutional level; and credible independent entities must be enlisted to assess and validate institutional claims for educational outcomes related to the public good need.

The role of technology and innovation, much of which is fostered by higher education institutions, is another high-visibility topic. Increasingly, people are concerned not just with the impact of innovation on productivity of firms in the commercial sector and on GDP, but also about its role in improving the public sector, sustainable growth, human and social welfare, health, civic engagement, and a host of nonmarket areas that mirror the scope of impacts for higher education. An ideal measure of the sector's productivity would capture non-pecuniary dimensions of outputs—choice in type of work, health, subjective well-being. But, of course, these are difficult valuation exercises as the stream of benefits are not transacted in markets and may accrue over very long periods.

3.3. Unit of measurement and level of aggregation

Individuals and departments

When productivity assessments at the individual level are used to measure performance, it is inevitable that the omission of other potentially relevant factors will lead to unsupportable conclusions.

In addition to the measurement complexities created by the joint production of multiple outputs, stakeholders and practitioners also are interested in productivity-related metrics aggregated at different levels, ranging from individuals (e.g., faculty) and academic departments at the most granular level to institutions, university systems, and the national level.[18] Productivity assessment at the individual level is a particularly thorny and controversial enterprise. Results are inevitably driven by the specification of the measure. For example, an English literature professor would appear to have a lower "productivity" than a finance professor when assessed based on graduates' salaries. Engineering departments, because they must invest in expensive equipment, would likely look worse than history departments in terms of costs per student hour. Dubious results will be obtained when any single criteria is disproportionately emphasized. An egregious example of such a misuse of performance concepts occurred when, in 2010, the office of the chancellor of Texas A&M University published what amounted to "a profit-and-loss statement for each faculty member, weighing annual salary against students taught, tuition generated, and research grants obtained" (*Wall Street Journal*, October 22, 2010). When such a simple metric is proposed as a tool for accountability or performance, it is inevitable that the omission of other potentially relevant factors will lead to unsupportable conclusions, the creation of perverse incentives, and a possibly damaging impact on the learning environment. In few circumstances would it make sense, for example, for a professor who teaches a small, specialized upper-level course as opposed to a very large general survey course, to be penalized in an assessment of performance.

18. The NAS report discusses productivity measures at different levels of aggregation, from individual educators, to departments, to institutions, to state systems, to aggregates in the National Income and Product Accounts.

Many of the same problems that plague individual level accountability schemes are also present at the department level. As in the English versus engineering teachers example above, different departments and universities will "score" differently along different criteria. A university's decision about whether to grow (or shrink) a department should not be driven exclusively by the prevailing wages in a profession dictating the cost of professors, the expected earnings of graduates, graduation rates, or any other singular criterion. Data are needed on all of these factors, and many more, if a holistic assessment approach is to be pursued by university decision makers.

Institutions

Interest in comparing institutions in terms of selectivity, quality of instruction, costs, earnings of graduates, and other dimensions is high. College rankings flourish on the web. The public (and, perhaps even more so, the media) has an insatiable appetite for easy-to-digest numbers which gloss over relevant pieces of information, such as that universities have strengths and weaknesses at the departmental level. Institution-level metrics also are of keen interest to policy makers. The demand for accountability initiatives designed to assess institutional performance is fully justified, especially in the case of public institutions, which obtain at least some of their funding from public budgets. Crucially, as discussed below, governments (most notably at the state level) have used quantitative value-added models to rate performance in a way designed to inform funding choices.

But measuring accountability quantitatively and qualitatively is a complicated business. What has made the nation's institutions great as a whole also makes them difficult to assess: "the diversity of colleges and universities is a huge asset to the nation, a factor any effective accountability system should take into account" (Carey, 2007, p. 27). When it comes to accountability, the major theme of this review is especially relevant: the question of interest should drive the measure, not the other way around. As eloquently conveyed by Carey: "Accountability systems begin with a conception of purpose: what an institution being held accountable is meant to be and do. Once that is established, the agents of accountability gather information about effectiveness…But it's not enough simply to gather information. Real accountability systems push institutions to act on that information in a manner that is designed to change what they do in order to make them more successful than they would otherwise be" (ibid).

An institution should be held accountable based on its purpose—**what it is meant to be and do.**

The first step, then, for developing accountability programs at the institution level is to specify the mission and then conceptualize what successful execution of that mission would look like. Next, the various technical measurement hurdles must be navigated. When comparing institutions, or tracking institutional performance over time, the selection of measures is central. Especially in value-added analyses, discussed in more detail in the next section, the measure must take into account students' academic baselines. If it does not, certain institutions will appear far more productive in terms of converting inputs (e.g., students) into outputs (e.g., graduates). Valuing degrees, whether based on learning, earnings, or some combination, requires controlling for characteristics (capabilities and backgrounds of students, i.e., the key input to the production process) that drive student choices and would be expected to attenuate the variation across institutions.

Institutions should only be compared with others in the same category and, even then, additional contextual information should be drawn upon, according to an NAS report.

The productivity model developed by the NAS panel—described in more detail below—requires assembling credit hour and degree information at the institution level (although there is nothing in principle that would make it less applicable to system or national level aggregations). Productivity, as the panel defined it, would factor in both credit hours completed and degrees awarded compared to labor costs and other expenses. By accounting for both degrees and credit hours, the model is designed to reward institutions for graduating students without penalizing them for having large numbers of part-time students. The NAS report made the point quite forcefully that institutions should only be compared with others in the same category (using something like the Carnegie Classification of Institutions of Higher Education) and, even then, additional contextual information should be drawn upon. The productivity of a research university cannot be compared to that of a liberal arts or community college, since they serve student population populations with different abilities, goals, and aspirations. The panel noted that institutional selectivity, program mix, size, and student demographics are among the most important contextual variables that must be controlled for when comparing productivity measures.

State systems

State policy makers have become increasingly interested in developing outcomes-based funding schemes for their public colleges and universities as a means to help achieve performance-oriented goals. According to the National Conference of State Legislators, 32 states[19] "have a funding formula or policy in place to allocate a portion of funding based on performance indicators such as course completion, time to degree, transfer rates, the number of degrees awarded, or the number of low-income and minority graduates." Federal agencies are also using or considering using quantitative measures of performance to create incentives for colleges to enroll students and provide access to higher education and to set funding levels (Dougherty et al., 2012).

As reported by Nate Johnson and Takeshi Yanagiura ("Early Results of Outcomes-Based Funding in Tennessee"[20]), "no state has done more than Tennessee to shift higher education funding toward outcomes. In other states, most of the core funding for higher education, including tuition and state appropriations, flows to colleges based on student enrollment."

19. The states are Arizona, Arkansas, Colorado, Florida, Illinois, Indiana, Kansas, Louisiana, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, New Mexico, New York, Nevada, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, South Dakota, Tennessee, Texas, Utah, Virginia, Washington, Wisconsin, and Wyoming (http://www.ncsl.org/research/education/performance-funding.aspx).

20. https://www.luminafoundation.org/resources/early-results-of-outcomes-based-funding-in-tennessee

For fiscal year 2011-12, Tennessee used an outcomes-based formula for the first time in its history. According to the formula, total outcomes-based funding was to account for 28 percent of funds (with tuition included). The impressive aspect of the plan is that it is not reliant on one performance measure. It includes measures indicating degree progress and completion (thereby creating incentives for institutions to go beyond focusing on enrollments and to help students successfully complete degree programs), but the formula also includes measures for "other important functions not directly related to degrees, such as research at research universities, and workforce development and dual enrollment at community colleges." The authors report that it is too early to assess the evidence from the policy, but that degree award data from Tennessee are "suggestively positive." Additional dimensions of performance can conceivably also be brought into play. A number of state longitudinal student databases (e.g., Washington, Virginia, Minnesota) are now linked to state earnings records, which will be useful for assessing productivity/outcomes.

National level

Policy makers and national accountants are interested in following employment, income, and other trends at the national level. They are concerned with measuring levels and change of output for the economy as a whole and for major sectors, including higher education. Generating such information is a central purpose of the National Income and Product Accounts from which key economic indicators such as GDP are generated. Historically, for a number of difficult-to-measure sectors, the value of output has been simply equated with the cost of inputs, making change in productivity impossible to detect. Accurate productivity measurement requires that the numerator (the outputs) and the denominator (the inputs) be estimated independently so that the sensitivity of former to changes in the latter can be detected. This input-output methodology has not been possible for the higher education because data have not supported independent measurement of the value of the sector's outputs.

The importance of higher education as a sector of the economy is evident in the aggregate figures. According to the Bureau of Labor Statistics,[21] in 2010, colleges and universities employed 3.6 million individuals, 2.6 million of those in professional positions. Based on the amount spent on inputs, the sector accounts (directly) for about 3.3 percent of gross domestic product (Soete et al., 2009), which makes it larger than a number of industries for which productivity data are routinely collected. It also accounted for about 10 percent of state budgets in recent fiscal years (National Association of State Budget Officers State Expenditure Report, 2011). Production of a large share of the nation's research and development by institutions of higher education accounted for $52 billion (60 percent of this funding derived from the federal government (National Science Board, 2010: 5-4).

Higher education is an important sector of the economy. In 2010:

3.6 million employed

3.3% of gross domestic product

About 10% of state budgets

21. See http://www.bls.gov/spotlight/2010/college/ which shows data on teacher salaries by field, earnings by graduates, and other information.

Several high profile efforts have improved the quality and usefulness of cost data, particularly at the level of the institution.

3.4. Productivity measurement in practice

Efforts are ongoing throughout all levels of higher education administration to quantify aspects of the production process, particularly as they relate to instruction. As emphasized above, costs are an essential aspect of performance measurement.[22] Several high profile efforts have improved the quality and usefulness of cost data, particularly at the level of the institution. The National Study of Instructional Costs and Productivity, often called the Delaware Cost Study, is a prime example. The program has been developed as a tool for benchmarking instructional costs, research, and public service expenditures at the academic discipline level. Participation has greatly increased in the program since its inception in 1992, with data now maintained for an annual average of 200 institutions. University administrators are able to access the data to establish benchmarks for "measuring faculty teaching loads, direct institutional costs, research expenditures, and separately budgeted scholarly activity against peer institutions."[23] These data provide valuable input for evaluations of the effectiveness of resource use compared with other institutions. By examining the experiences of like institutions, the Cost Study has been used as a planning resource to project costs associated with developing and implementing new programs.

At the sub-institution level, course-based or activity-based costing approaches are also being developed that provide detailed revenue and cost data.[24] Activity-based costing is a methodology designed to "identify activities in an organization and assign the cost of each activity with resources to all products and services according to the actual consumption by each" (Anguiano, 2013). William Massy—writing for the TIAA Institute Higher Education Series: Understanding Academic Productivity, of which this review is part—reports on use by the office of the provost at the University of California, Riverside to implement an enhanced activity-based costing model and software tool at the course level. The objective of the tool is to report "the activities, costs, and revenues associated with the course portfolio and other university functions, which gives both academic and financial decision-makers better insight into their planning and budgeting options. For example, chairs and deans can do better in choosing the most appropriate mix of instructional models given their resource constraints, and demonstrate the efficacy of their choices." In this type of application, activity-based costing is capable of questions such as "*What is the best way to deploy resources (people and funding) to achieve our educational mission? What is the best way to achieve any given curriculum within resource constraints?* [and] *Could a different allocation of resources achieve better results with the same investment of time and money?*" (Massy, 2016, p. 2).

22. It is worth noting again that, from the viewpoint of consumers of higher education—students and their parents—the issue of who pays the costs is as important as the level of costs. Large tuition increases at public universities in recent years have resulted more from declining state appropriations than from increasing costs of production. Without adequate state investment of taxpayer dollars, institutions will either continue to become more expensive to students, or they will have to sacrifice quality.

23. http://www.udel.edu/IR/cost/consortial.html

24. This kind of costing exercise can be carried out at the departmental or institutional level.

While activity-based costing confronts the problem that colleges and universities do not always fully understand the cost of their educational activities, such tools address only one of the missing components of productivity measurement. The inadequacy of information essential for making efficient cost reduction or other spending decisions is rooted even more fundamentally to the missing capacity to relate costs to outcomes. Because costs are now well understood, recent research has turned toward the more difficult task of measuring outputs and linking inputs to outputs.

Recent research has turned toward the more difficult task of **measuring outputs and linking inputs to outputs.**

Scorecards, dashboards, and rankings

There has always been keen interest in metrics for gauging institutional performance—on the basis of student access, completions, costs, and outcomes after graduation—in a way that informs student choices. As evidenced by a spate of recent initiatives, this demand for useful metrics is intensifying as recognition solidifies that sound data are essential to understanding and improving educational outcomes. A recent paper by Jennifer Engle (*Answering the Call: Institutions and States Lead the Way Toward Better Measures of Postsecondary Performance*[25]) outlines the goals of an initiative by the Gates Foundation seeking to lead the way in developing a rich data platform for the sector. Goal 1 of the initiative is to improve data and metrics over those that are commonly produced and that tend to focus on traditional students while ignoring the "new normal in higher education, which often includes students attending college—or colleges—in new ways en route to their credentials. Colleges and universities, and the data systems that support them, must adjust to and reflect the experiences and outcomes of all students, not just the outdated 'traditional' student profile" (Engle, p. 1). The data framework is to be based on a consensus among institutions, organizations, and states that have implemented their own data collection efforts. Once data sources and metrics have been improved, the foundation will document the progress that institutions and states have made through the use of improved data, and support adoption and use of these metrics.

The stated goal of the Gates Foundation initiative, then, is to provide information that can be used in student decision-making as well as by institutions and system leaders to support improvement in their programs and communities; this requires both making existing data sets more publicly available and developing new and improved metrics. The metrics also will be used to evaluate the impact of the foundation's own investments in higher education that are geared toward increasing student attainment of career-relevant credentials and closing attainment gaps across socioeconomic groups. An ultimate goal of the foundation's initiative is to improve the quality and relevance of postsecondary data across the field in a way that can "better inform higher education practice and policy decisions that, in turn, can boost college access and success across the country" (Engle, p. 1).

25. http://postsecondary.gatesfoundation.org/wp-content/uploads/2016/02/AnsweringtheCall.pdf

Governments at various levels also have ramped up efforts to improve data on the performance of higher education. The U.S. College Scorecard, recently redesigned by the Obama administration, has as its goal to “empower students to choose the college that is right for them.”[26] The federal government collects data on a wide range of the sector’s activities—but it has typically shared just 20 to 30 data fields out of a total of 1,300 that it collects. Clearly, there is value to stakeholders in making these data more accessible; with adequate protections in place, it can be done with little risk to privacy and confidentiality. The goal of the College Scorecard is to provide greater visibility to the neglected side of the cost-benefit calculus—student outcomes—and to shift away from the subjective emphasis on prestige, nurtured by the rise of college rankings, and toward factors more closely linked to the educational experience and benefits that accrue to students. Toward that end, the database makes public institution-level data on colleges’ graduation rates, students’ median earnings 10 years after entering the school,[27] and the percentage of students paying back their college loans.

While the College Scorecard adds rich detail on more than 7,000 U.S. higher-education institutions, which will be a useful resource to college applicants and others, it is still inherently incomplete (as any one assessment tool is destined to be). This is not a criticism of the effort; rather, it is a function of the underlying analytic question. Predicting the impact that a particular college will have on a graduate’s earnings—a major purpose of the scorecard—is fraught with measurement traps. It is a classic case of being able to establish correlation but not causation. James Stewart, writing for the *New York Times,*[28] points out that of course “graduates of the Massachusetts Institute of Technology (average postgraduate earnings $91,600, according to the Scorecard) and Harvard ($87,200) do well. That’s because the students they admit have some of the highest test scores and high school grade point averages in the country, reflecting high intelligence and a strong work ethic—two factors that cause high future earnings. That is generally true regardless of where such students attend college, as long as they go to a reputable four-year institution, various studies have shown.” In other words, because they are highly selective, it is very difficult to disentangle high earnings in a way that convincingly attributes them to these institutions.

26. https://www.whitehouse.gov/the-press-office/2015/09/12/fact-sheet-empowering-students-choose-college-right-them

27. This requires data from tax returns, which are only available for students who received federal loans or grants, which therefore means the sample is truncated to exclude most students from high-income families.

28. October 1, 2015, http://www.nytimes.com/2015/10/02/business/new-college-rankings-dont-show-how-alma-mater-affects-earnings.html?_r=0 [accessed April 2016]

Li Zhou, writing for the *Atlantic,*[29] concludes that, while the U.S. College Scorecard is a step forward, it is by design limited to what is readily measureable, which omits many of the private and social benefits of higher education: “If you go to M.I.T. and earn a degree in engineering, you’re going to make more than if you go to Oberlin and major in music performance,” Professor Muller said. “But you already know this. To rank the value of colleges based on the ultimate earnings of their graduates radically narrows the concept of what college is supposed to be for.” The scorecard is meant to be a tool for use by potential consumers of higher education in combination with other information interpreted with consideration of individual-specific factors. It is not intended as a tool for administrators and legislators for accountability purposes in funding decisions; however, as has been repeatedly demonstrated, high-visibility tools do infiltrate into these processes to create incentives that affect the direction of change.[30] Responsible curators of any metric must be cognizant that institutional behavior responds to the incentives embedded within measurement and should take steps to maximize the likelihood that measured performance is the result of real success rather than manipulative behaviors (NAS, 2012).

The U.S. College Scorecard is a tool intended to be used in combination with other information, interpreted with consideration of individual-specific factors.

Institutions themselves also are compelled to publicize dashboards of information in order to internally assess progress on strategic priorities, and to respond to external calls for greater accountability from policy makers and the public. Terkla (2011), associate provost for institutional research and evaluation at Tufts University, endeavored to catalogue the use of public dashboards with their indicators and found 66 institutions ranging from small colleges to major research universities that compiled them. She found little commonality of the indicators used for the dashboards, which supports the notion “that institutions develop their indicators based on their specific strategic plan and institutional characteristics,” and thus they appear to be mainly for internal, not external, use (Terkla, p. 1). Table 1 shows the most common indicators by frequency of use grouped in 11 broad categories.

29. Sept 15, 2015, http://www.theatlantic.com/education/archive/2015/09/obamas-new-college-scorecard-flips-the-focus-of-rankings/405379/ [accessed April 2016].

30. Grewal, Dearden, and Lilien (2008) document ways in which universities strategically deploy resources in an attempt to maximize their rankings in U.S. News & World Report and elsewhere. Avery, Fairbanks, and Zeckhauser (2003) and Ehrenberg and Monks (1999) find that the ranking system distorts university admissions and financial aid decisions.

Table 1: The Most Common Performance Indicators (by percentage using them) for Institutions and Their Boards

Category	Indicator Group	Number of Dashboards Using (N=66)	Percent
Financial	Endowment & Expense Data	53	**80.3**
	Advancement	48	72.7
	Financial Aid Figures	42	63.6
	Fees/Tuition Data	31	47.0
Admissions	Admissions Scores	52	**78.8**
	General Admissions Data	47	71.2
	Graduate Admissions	14	21.2
Enrollment	Enrollment Figures	51	**77.3**
	Enrollment Figures (Special Population)	47	71.2
Faculty	Faculty–General	51	**77.3**
	Faculty Composition–Special Population	22	33.3
Student Outcomes	Graduation Rates	48	**72.7**
	Retention Rate	47	71.2
	Measures of Success	27	40.9
	Enrollment Awards	15	22.7
	Graduation Rates–Special Population	10	15.2
Student Engagement	Student Body–Engagement	38	**57.6**
Academics	Student/Faculty contact	36	**54.5**
	Academic Information	31	47.0
Physical Plant	Physical Plant	23	**37.9**
Satisfaction	Student Satisfaction	23	**34.8**
	Employer/Staff, Other Satisfaction	7	**10.6**
	Faculty Satisfaction	3	4.5
Research	Research	23	**34.8**
External Ratings	Peer Assessment Data	14	**21.2**

Source: Terkla (2011).

A system-wide effort to launch a "Student Success Scorecard" was launched by the California Community Colleges Chancellor's Office in March of 2013. The scorecard provides data on California's 112 community colleges and 2.4 million students and including data on graduation, retention and transfer rates for each of the colleges, and for the overall system.

All of these approaches, from the national-level College Scorecard to the institution-level products, attempt to offer statistics that will be useful to stakeholders by including concepts of quality that are relevant to performance and productivity in higher education. As the range of methodologies shows, there is no commonly accepted method for measuring the quality of a college education, and this remains a major challenge. Jenkins and Rodríguez (2013) offer a detailed discussion of metrics for tracking quality of outputs and outcomes with the goal of "Improving Productivity in Broad-Access Postsecondary Institutions." The authors observe that, in the past decade or so, efforts have accelerated to measure the quality of the *outputs* of undergraduate education. They point to four categories of indicators, which while they cannot provide a comprehensive measure, have proved to be useful pieces of the information puzzle for assessing productivity (p. 191):

No commonly accepted method for measuring the quality of a college education exists.

Accelerating efforts to measure the quality of the outputs of undergraduate education are pointing to four categories of indicators.

Standardized tests. Test scores are designed to indicate demonstrated knowledge or analytic ability—and ideally to show changes (presumably improvements) at one point in time relative to an earlier point. Some researchers (e.g., NAS, 2012) have proposed that a uniform college exit exam be established so that institutions can be compared across a standardized dimension. Some colleges already use instruments such as the Collegiate Learning Assessment to measure how much students are learning. The trick, if one is to measure learning during college, is that a baseline has to be created to measure students' levels of knowledge when they started college (see the discussion on value-added models below). These are, by their nature, individual level measure, but averages may be used to compare groups of students and institutions. Liu (2011) compares the results of two different methodologies that measure value added of institutions using standardized test scores.

External certification. Some occupational fields (e.g., accounting and nursing) maintain certification or licensure assessment systems established by industry or by professional organizations, which provide indicators of quality.

Learning outcomes standards. Specification of the knowledge content and skills in which students are expected to be proficient are a central element of the higher education accreditation process. A 2009 survey of college leaders by the National Institute for Learning Outcomes Assessment found that many undergraduate institutions have not fully adopted such assessments; community colleges are more likely than research universities to have done so.

Earnings of graduates. As described above, post-graduation earnings are a quantifiable measure of the economic benefits of a college degree that have a relationship (even if a murky one) to the quality of an education received. Life-cycle earnings have been the primary outcome of interest among labor economists whose research tends to focus on the pecuniary, market returns to education.

Due to concerns among policy makers that the U.S. is falling behind other industrialized nations in producing a workforce with adequate science and technical skills, a particular focus in recent years has been placed on measuring the quality of undergraduate education in science, technology, engineering, and mathematics (STEM) fields. The National Science Board's Science and Engineering Indicators, the California State University System's STEM Dashboard, and regional accreditations of undergraduate education offer examples of various approaches to measuring equity in undergraduate STEM. Led by the National Science Foundation, proposals have been forwarded to identify objectives for developing national indicators of quality in two-year and four-year STEM education. The agency's Improving Undergraduate STEM Education initiative includes a "foundation-wide integrated framework for the agency's investments in undergraduate STEM education." NSF's investments in undergraduate STEM education are coordinated through this initiative to enhance coherence and impact and to use shared metrics and evaluation approaches.[31]

Value-added models

A major avenue through which considerations of quality have entered into productivity measurement discussions is the value-added literature. Value-added modeling sometimes is pursued to address pressures on institutions to demonstrate a return on investment by reporting the value they generate to students and their families, both financially and terms of other factors—e.g., providing students with an intellectually stimulating environment, proximity to amenities, quality of life, health care, etc. Carefully constructed value-added measures isolate the contribution to student outcomes of completing college (most often, in such studies, limited to the earnings component) over what would be predicted to have happened, based on student characteristic, without college. The calculation is not exactly a measure of return on investment, but rather a way to compare colleges on a level playing field by "adjusting outcome measures for the relative advantages or disadvantages faced by diverse students pursuing different levels of study across different local economies" (Rothwell and Kulkarni, 2015).

The main methodological hurdle that must be overcome in value-added modeling is "unobserved selection."

Student outcomes are affected by students' own characteristics, such as ability and work ethic.

As identified above, the main methodological hurdle that must be overcome in value-added modeling is "unobserved selection." Unobserved selection occurs because student outcomes are affected by students' own characteristics, such as ability and work ethic (variables that are not typically captured in standard administrative data sources), that systematically draw them to different institutions. This pattern of self-selection makes it

31. For documentation of the NSF program, go to http://www.nsf.gov/about/budget/fy2017/pdf/43_fy2017.pdf

difficult to isolate the portion of observable student achievement that is attributable to attending a given college from the portion attributable to other factors. Dale and Krueger (2002) analyzed the monetary return to attending highly selective colleges using individual-level information on schools applied to and accepted in an effort to minimize the influence of unobserved selection. They found that: "the return to college selectivity is sizeable for both cohorts in regression models that control for variables commonly observed by researchers, such as student high school GPA and SAT scores." However, after adjusting for unobserved student ability by controlling for the average SAT score of the colleges that students applied to, their "estimates of the return to college selectivity fall substantially and are generally indistinguishable from zero" (p. 1491).[32]

Cunha and Miller (2014) produced value-added estimates for individual institutions as well. Their method for measuring value added—described as a practical application of the Dale and Krueger (2002) methodology—controls for a set of applications and acceptances designed to mitigate unobserved selection into schools and involves the use of administrative data now becoming available in some states. Using data on 30 public universities from the state of Texas, the researchers are able to follow the universe of students, from application, through public university, and into the labor market with the goal of assessing the relative effectiveness of different institutions, as prescribed by the Spellings commission report (2006). Cunha and Miller's findings (p. 64) are quite consistent with those of Dale and Krueger:

> In specifications that do not control for selection, we find large, significant differences across colleges in terms of persistence, graduation, and earnings; however, these differences decrease substantially when we control for selection. In light of the growing interest in using value-added measures in higher education for both funding and incentivizing purposes, our methodology offers unique evidence and lessons for policy makers.

In light of the growing interest in using value-added measures for both funding and incentivizing purposes, **Cunha and Miller's methodology offers unique evidence and lessons for policy makers.**

Rothwell and Kulkarni (2015) use alumni earnings data, measured six years after admission, within a value-added approach updated with College Scorecard data to rank several thousand two- and four-year colleges based on graduates' economic outcomes.[33] As with the studies cited above, the authors proceed under the assumption that future earnings of alumni are affected by student characteristics (such as their academic preparation, age, racial or ethnic background, and family income), the type of college (a community college or research university, for example), the location of the college (as in a big city with many jobs compared to a small town), and the qualities of the college. To estimate the college's contribution to

32. The authors did find notable exceptions for certain subgroups: "For black and Hispanic students and for students who come from less-educated families (in terms of their parents' education), the estimates of the return to college selectivity remain large, even in models that adjust for unobserved student characteristics."

33. http://www.brookings.edu/research/reports2/2015/10/29-earnings-data-college-scorecard-rothwell

student earnings, earnings for an individual college are predicted based on these attributes and compared to actual outcomes. The value-added of the college can be thought of as the difference between expected and actual outcomes, or the sum of unmeasured and measured qualities of the college. Among the findings from the Rothwell and Kulkarni analysis was that graduates of some colleges enjoy much more economic success than their characteristics at time of admission would predict. The key factors creating high value added in terms of the financial success of an institution's alumni were found to be: training for high-paying careers in technical subjects such as in those with a medical or engineering orientation; the share of graduates prepared to work in STEM occupations; the percentage of students finishing their award within at most 1.5 times the normal time (three years for a two-year college, six years for a four-year college); and the average monthly compensation of all teaching staff. The study found the value-added measures to be fairly reliable over the period of analysis, from the cohort entering in 1997–1998 to the cohort entering in 2005–2006, but that several colleges had made large improvements on value-added over that period, "including Stanford, Yale, Georgetown, and Emory, as well as Glenville State College and Alaska Pacific University."

Productivity improvement is vital to maintaining a high-quality education sector and to making college affordable for as many people as possible.

As with all measures, the usefulness of value-added models depends on the question being asked. They are valuable for informing students and administrators about the prospects of a degree or other credential from a given institution to contribute to students' economic advancement. As pointed out by Rothwell and Kulkarni, most value-added measures used in college quality research do not tell us much about student improvement that manifests as learning gains. At the macro level, such models may provide convincing estimates of the average wage premium paid to college graduates relative to those with less education.

Putting inputs and outputs together: measuring productivity

In an environment defined largely by rigid resource constraints, productivity improvement—increasing college enrollment and the number of graduates, as well as the amount of learning and innovation relative to the inputs used—is seen as vital to maintaining a high-quality education sector and to making a college education affordable to as much of the population as possible. Under this premise the NAS (2012, p. 2) panel was charged with identifying an analytically well-defined concept of productivity for higher education; recommending practical guidelines for its measurement; and constructing valid productivity measures to "supplement the body of information used to (1) guide resource allocation decisions at the system, state, and national levels and to assist policy makers who must assess investments in higher education against other compelling demands on scarce resources; (2) provide administrators with better tools for improving their institutions' performance; and (3) inform individual consumers and communities to whom colleges and universities are ultimately accountable for private and public investments in higher education."

In the report, the base-line productivity measure of the instructional component of higher education was estimated as: "the ratio of (a) changes in the quantity of output, expressed to capture both degrees (or other markers of successful completion) and passed credit hours to (b) changes in the quantity of inputs, expressed to capture both labor and non-labor factors of production" (p. 3). The assumption embedded in the numerator reflects thinking in the economics literature on human capital (e.g., Bailey et al., 2004; Barro and Lee, 2010) that "education adds to a student's knowledge and skill base, even if it does not result in a degree." An example of the productivity calculation was provided for an unnamed institution using the Integrated Postsecondary Education Data System (IPEDS) maintained by the National Center for Education Statistics. The sponsor of the study, the Lumina Foundation, later used the methodology to calculate productivity estimates for a range of institutions, but those estimates have not (to this author's knowledge) been reproduced in a published study. This is probably a good thing given that the authors explicitly stated that estimates derived using their productivity measure were not yet ready to be used in accountability assessments or operational policy making. The reason for the cautious tone was that the productivity measure did not explicitly take into account the all-important variation or changes in the quality of outputs or inputs.[34]

In the final analysis, and by the authors' own admission, the experimental measure developed in the report did not directly advance all of these objectives, but it did push the discussion forward and offer a conceptual framework for future work. Calling attention to the need for improved productivity metrics and for their implementation in the broader performance assessment picture, the panel concluded that: (1) productivity should be a central part of the higher education conversation; (2) conversations about the sector's performance will lack coherence in the absence of a well-vetted and agreed-upon set of metrics, among which productivity is essential; (3) quality should always be a core part of productivity conversations, even when it cannot be fully captured by the metrics; and (4) the inevitable presence of difficult-to-quantify elements in a measure should not be used as an excuse to ignore those elements.

Continued work on value-added models is showing the way toward producing far more comprehensive measures than were possible in the recent past.

To a large extent, productivity and performance research is still missing the quality piece, which must be evaluated alongside the quantity data. However, continued work on value-added models and other measures described above is now closing the gap and showing the way toward producing far more comprehensive measures of higher education than were possible in the recent past. How to weight the different components in an analysis (for example, valuations of the output in terms of graduates' learning gains versus their economic gains, or the valuation of private versus public benefits) will always require subjectivity and subject matter expertise, and will depend on the specific question being asked. Part of the attraction of a multidimensional productivity measure that incorporates

34. It is worth noting that one indirect effect is captured by the NAS panel's productivity equation to the extent that higher quality inputs lead to higher graduation rates, a key component of the numerator. The example is given that "if small classes or better teaching (inputs of different quality) lead to higher graduation rates, this will figure in the output total as a greater sheepskin effect—that is, an added value assigned for degree completion" (p. S-4).

both the input (cost) and output (benefit) sides of the process is that it highlights the problem with single, high-stakes measures of performance that make gaming the system simpler and that a range of measures almost always will be preferable for the purpose of weighing the overall performance of a department, and institution, or the sector at large.

4. Discussion

Data Issues

NAS (2012), Massy et al. (2012), Rothwell and Kulkarni (2015), Engle (2016), and others have identified several specific examples of the kinds of data that need to be developed or better exploited in order to embark upon the next phase of productivity measurement.[35] Some improvements can be made by supplementing existing data sources. As recommended by the NAS (2012) panel, the National Center for Education Statistics should "examine the feasibility of (a) requesting modifications to university accounting systems and then updating the Integrated Postsecondary Education Data System (IPEDS) to identify full-time employees by labor category...and (b) calculating total compensation for each category and function." The idea behind this kind of added detail would be to make comparison of the cost of these activities by institution—crucial to quantifying inputs in the productivity measure—more feasible. The capacity to track all course hours that go into the production of degrees also would be enhanced if students' credit hours could be linked to degree or field. To move in this direction, NAS (2012) recommended that "institutions should collect credit-hour data in a way that follows students, and not only the departments that teach them." At many institutions, the necessary information already exists in administrative data sources. In turn, IPEDS could compile and report these data along with the numbers of degrees awarded.

Detailed productivity measurement requires additional information to estimate output.

Detailed productivity measurement requires other kinds of information in order to estimate output (and outcomes) as well. In order to better calculate graduation rates and estimate the cost and value of degrees, comprehensive longitudinal student databases and more accessible administrative sources are needed. The potential of administrative data sources—maintained at various levels, ranging from institutions' accounting and management systems to those of the federal statistical agencies—depends heavily on the ability of researchers and policy analysts to link records across state boundaries and across elementary, secondary, postsecondary, and workforce boundaries (Prescott and Ewell, 2009).

35. NAS (2012) appendix C contains a detailed list of data sources on education statistics. Data sources are categorized by the unit of analysis or their collecting agencies. Unit of analysis may be an institution, faculty, student, or household. Collection agencies include federal agencies, state agencies, or private data-collection agencies.

In the same vein, outcome measurement has been enhanced by the capacity to link student level administrative data with longitudinal data. A number of State Longitudinal Data Systems (SLDSs) provide rich sources of data that can be linked over time and to additional data sources for purposes of administration, reporting, research, and to inform policy and practice. The Census Bureau's Longitudinal Employer-Household Dynamics program combines wage data from state agencies with demographic and other information from the Census Bureau. The Bureau of Labor Statistics continues to support programs that facilitate multi-state links of unemployment insurance wage records and education data, which creates new opportunities for research estimating the return on investment from postsecondary training and on placement rates in various occupations. Such sources are crucial for researching student outcomes as they progress into the labor force (Snyder, Dillow, and Hoffman, 2007).[36]

Rothwell and Kulkarni (2015) note data limitations of scorecards and other sources that are constraining progress in value-added modeling, including: (1) incomplete coverage—for example, the College Scorecard data are limited to federal aid recipients, a nonrandom sample that biases estimates of earnings, for example; (2) lack of outcomes data disaggregated by field of study—this is true for all multi-state databases, although disaggregated earnings are reported by Texas Consumer Resource for Education and Workforce Statistics for alumni who remain in the state of Texas; and (3) few measures of learning exist in the scorecards—"while achieving higher earnings are often a secondary goal, acquiring a body of useful or meaningful knowledge is perhaps the chief goal of higher education."

Final thoughts

Thanks to an increasingly rich literature on the topic, our conceptual understanding of productivity measurement, applied to complex sectors such as higher education, has evolved markedly in recent decades. Without minimizing the remaining conceptual challenges, it can be reasonably asserted that a point has been reached where further progress in measuring the activities of the sector—in a way that informs policy, operational decision-making, and stakeholder concerns—rests on continued improvement of source data. As stated by Engle (2016) in the previously mentioned Gates Foundation report: "Better data alone will not guarantee better student outcomes, but a lack of better data will guarantee that our efforts to improve those outcomes will fall short of their potential." Work is ongoing on numerous fronts to realize this goal for the higher education sector and, as this paper indicates, these developments are highly encouraging.

"Better data alone will not guarantee better student outcomes, but a lack of better data will guarantee that our efforts to improve those outcomes will fall short of their potential."

(Engle, Gates Foundation Report, 2016)

36. Of course there are difficulties to overcome. For example, the capacity to follow students becomes problematic when students leave state for college or work (Cunha and Miller).

More coordination of effort is needed to produce a systematic framework for addressing a wide range of issues.

Regardless of how far the statistics and the data from which they are constructed in advance in terms of quality, analyst expertise and interpretation always will play an essential role in their interpretation and in taking into account their limitations. This will continue to be the case particularly for unidimensional measures often used as proxies for productivity and performance out of convenience due to their ready availability. Graduation rate, unit cost, time to degree, earnings data, and similar metrics have utility in comparing institutions and programs and for monitoring their performance over time, but such data must be used cautiously and with knowledge of their drawbacks. Experience has shown that casual and indiscriminate application to policy and operational questions can lead to distorted conclusions and the creation of perverse incentives.

It is also clear that, because so many topics of interest are in play, no data source will accommodate all or even most purposes. Without question, more coordination of effort is needed to produce a systematic framework for addressing a wide range of issues of importance to various stakeholders. Therefore, the capacity to link administrative and survey data in creative ways to longitudinally track the mix of inputs, costs, and student performance and outcomes is a key requirement to progress. In many ways, the concept of productivity provides that framework by organizing the process into inputs and outputs, and their many subcomponents. As stated in NAS (pp, 16-17) "the idea that instructional productivity may potentially be increased by altering the way inputs in the production function are combined highlights why improved measurement is so important. Potential improvement in productivity also justifies requirements that colleges and universities systematize collection of data on expenditures and the volume and quality of inputs and outputs. Routine generation and collection of such data is a prerequisite for wider efforts to improve productivity and enable external stakeholders to hold institutions accountable."

About the Author

Christopher Mackie

Christopher Mackie is a program officer specializing in economic measurement and statistics with the Committee on National Statistics, National Academies of Sciences, Engineering, and Medicine. He has served as study director for a range of projects, including those that produced the reports *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes; Beyond the Market: Designing Nonmarket Accounts for the United States; Understanding Business Dynamics: an Integrated Data System for America's Future; Accounting for Health and Health Care: Approaches to Measuring the Sources and Costs of Their Improvement; Improving Measurement of Productivity in Higher Education; and Subjective Well-being: Measuring Happiness, Suffering, and Other Dimensions of Experience.* He is co-editor of a major upcoming National Academies volume on the labor market and fiscal impact of immigration. He is also author of Canonizing Economic Theory: How Theories and Ideas Are Selected in Economics. He holds a Ph.D. in economics from the University of North Carolina.

References

Anguiano, M. (2013). *Cost Structure of Post-Secondary Education: Guide to Making Activity-Based Costing Meaningful and Practical.* Virginia Beach, VA: Pilbara Group. https://www.ucr.edu/about/admin/docs/Cost-Structure-of-Post-Secondary-Education-Guide-v2.pdf

Archibald, R.B., and Feldman, D.H. (2011). *Why Does College Cost So Much?* Oxford, England: Oxford University Press.

Arum, R., and J. Roksa (2011). *Academically Adrift: Limited Learning on College Campuses,* University of Chicago Press.

Babcock, P., and M. Marks (2011). "The Falling Time Cost of College: Evidence from Half a Century of Time Use Data," *The Review of Economics and Statistics*, MIT Press, vol. 93(2), pages 468-478, December.

Bailey, T., Kienzl, G., and Marcott, D.E. (2004). *The Return to a Sub-Baccalaureate Education: The Effects of Schooling, Credentials and Program of Study on Economic Outcomes*. Washington, DC: U.S. Department of Education.

Barro, R.J. (1991). "Economic Growth in a Cross-Section of Countries," *Quarterly Journal of Economics.* 106, pp. 407-43.

Barro, R.J., and Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics, 104*, 184-198. doi: http://dx.doi.org/10.1016/j.jdeveco.2012.10.001

Barrow, C.W. (1990). *Universities and the Capitalist State: Corporate Liberalism and the Reconstruction of American Higher Education, 1894 – 1928*. Madison: University of Wisconsin Press.

Baumol, W.J., and Bowen, W.G. (1966). *Performing Arts: The Economic Dilemma.* Cambridge, MA: MIT Press.

Belfield, C. (2012). "Measuring Efficiency in the Community College Sector," Working Paper 43 (New York: Community College Research Center, Teachers College, Columbia University, April.

Bosworth, B. (2006), "Productivity in Education and the Growing Gap with Service Industries," https://net.educause.edu/ir/library/pdf/ffpiu044.pdf

Bowen, W.G., Chingos, M.M., and McPherson, M.S. (2009). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press.

Brinkman, P. (1985). "Instructional costs per student credit hour: Differences by level of instruction," Boulder, CO: National Center for Higher Education Management Systems.

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

Bruni, F. (2015), "College's Priceless Value," *The New York Times,* Feb. 11, 2015.

Carey, K. (2007). “Truth without action: The myth of higher-education accountability,” *Change, 39*(5), 24-29, September/October.

Cooke, M.L. (1910). *Academic and Industrial Efficiency: A Report to the Carnegie Foundation for the Advancement of Teaching*. New York: Carnegie Foundation for the Advancement of Teaching.

Cunha J.M., and T. Miller (2014). “Measuring value-added in higher education: Possibilities and limitations in the use of administrative data,” Economics of Education Review 42, 64-77

Cutler, D. (2004). *Your Money or Your Life: Strong Medicine for America's Health Care System*. New York, NY: Oxford University.

Dale, S.B., and A.B. Krueger (2002). “Estimating the Payoff to Attending a More Selective College: an Application of Selection on Observables and Unobservables,” *Quarterly Journal of Economics*, 117(4), pp. 1491–1528.

Desrochers, D.M. (2011). *Estimating the production cost of degrees by level and discipline using national data: An exploration of methodologies.* Unpublished manuscript.

Dundar, H., and D.R. Lewis (1998), “The Determinants of Research Productivity in Higher Education,” *Research in Higher Education,* Vol. 39, No. 6, pp. 607-631.

Dougherty, K.J., R.S. Natow, and B.E. Vega (2012). “Popular but unstable: Explaining why state performance funding systems in the United States often do not persist,” *Teachers College Record*, 114 (3), pp. 1–42.

Ehrenberg, R.G. (2012). “American higher education in transition,” *Journal of Economic Perspectives 26*(1):193-216.

Engle, J. (2016). *Answering the Call: Institutions and States Lead the Way Toward Better Measures of Postsecondary Performance*, Report of the Bill & Melinda Gates Foundation.

Farrell, M. (1957). “The measurement of productive efficiency,” *Journal of the Royal Statistical Society, Series A*, 120, pp. 253-281.

Garrett, T., and W. Poole (2006). “Stop Paying More for Less: Ways to Boost Productivity in Higher Education,” *Regional Economist* (Federal Reserve Bank of St. Louis, January). https://www.stlouisfed.org/publications/regional-economist/january-2006/stop-paying-more-for-less-ways-to-boost-productivity-in-higher-education

Gates, S., and A. Stone (1997). “Understanding Productivity in Education,” Institute on Education and Training, RAND Corp., January.

Griliches, Z. (1987). “Productivity: measurement problems,” In J. Eatwell, M. Milgate and P. Newman, eds. *The new Palgrave dictionary of economics, vol 3.* New York, USA, Stockson Press.

Harnisch, T.L. (2011). Performance-based funding: A re-emerging strategy in public higher education financing, American Association of State Colleges and Universities, Policy Matters.

Harris, D., and Goldrick-Rab, S. (2011). *Higher Education in Wisconsin: A 21st Century Status Report.* Madison: Wisconsin Covenant Foundation.

Jenkins, D., and Rodriguez, O. (2013). Access and Success with Less: Improving Productivity in Broad-Access Postsecondary Institutions. The Future of Children Spring; 23(1):187-209.

Johnson, N. (2009). *What does a college degree cost?* A Delta Project White Paper. Washington, DC: Delta Project on Postsecondary Costs, Productivity, and Accountability. Retrieved from http://www.deltacostproject.org/resources/pdf/johnson3-09_WP.pdf

Johnson, N. (2016). "State Postsecondary Education Productivity and the Limits of Institutional Measurement," TIAA Institute Higher Education Series: Understanding Academic Productivity.

Kokkelenberg, E.C., Sinha, E., Porter, J.D., and Blose, G.L. (2008). *The Efficiency of Private Universities as Measured by Graduation Rates*. Working Paper No. 113. Ithaca, NY: Cornell Higher Education Research Institute.

Liu, O. (2011). "Value-added assessment in higher education: A comparison of two methods," *Higher Education*, 61, pp. 445–461.

Lucas, R. E., Jr. (1988). "On the mechanics of economic development," *Journal of Monetary Economics,* 22(June):3-43.

Mankiw, N. G., D. Romer, and D. N. Weil. (1992). "A contribution to the empirics of economic growth," *Quarterly Journal of Economics,* 107(May):407-437.

Massy, W.F. (2016). *Course-Level Activity-Based Costing as an Academic and Financial Tool*. TIAA Institute Higher Education Series: Understanding Academic Productivity.

Massy, W.F., T.A. Sullivan, and C. Mackie (2012). "Data Needed for Improving Productivity Measurement in Higher Education," *Research & Practice in Assessment*, volume 7, Winter.

Moretti, E. (2004). "Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data," *Journal of Econometrics*, 121, pp. 175–212.

National Science Board (2010). *Science and Engineering Indicators 2010.* Arlington, VA: U.S. National Science Foundation.

OECD. (2015), *Education at a Glance 2015: OECD Indicators*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/eag-2015-en

Oreopoulos, P., and Salvanes, K. G. (2011]). "Priceless: The nonpecuniary benefits of schooling," *The Journal of Economic Perspectives*, 25(1), 159–184.

Porter, S., and Toutkoushian, R. (2006). "Institutional research productivity and the connection to average student quality and overall reputation," *Economics of Education Review*, *25*, 605-617.

Ramsden P., and I. Moses (1992). "Associations between research and teaching in Australian Higher Education," *Higher Education*, 23 (3) (1992), pp. 273–295.

Romer, P. M. (1990) "Human Capital and Growth. Theory and Evidence," *Carnegie-Rochester Conference Series on Public Policy.*

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

Shedd, J. (2003). "The history of the student credit hour, "*New Directions for Higher Education 122*:5-19.

Snyder, T.D., Dillow, S.A., and Hoffman, C.M. (2007). *Digest of Education Statistics 2006* (NCES 2007-017). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S. Government Printing Office.

Soete, L., Guy, K., and Praest Knudsen, M. (2009). *The Role of Community Research Policy in the Knowledge-Based Economy*. Expert Group Report to DG Research, European Commission. Luxembourg: Publications Office of the European Union. Available: http://ec.europa.eu/research/era/pdf/community_research_policy_role.pdf

M. Spellings (2006), *A test of leadership: Charting the future of U.S. higher education*. U.S. Department of Education Report.

Sullivan, T. A., Mackie, C., Massy, W. F., and Sinha, E. (Eds.). *Improving measurement of productivity in higher education: Panel on measuring higher education productivity: Conceptual framework and data needs*. Washington, DC: The National Academies Press.

Terkla, D. (2011), "The Most Common Performance Indicators for Institutions and Their Boards," Trusteeship Magazine, *Association of Governing Boards of Universities and Colleges*, January/February.

Tilak J.B.G. (2008), "Higher Education: A Public Good or a Commodity for Trade? Commitment *to* Higher Education or Commitment *of* Higher Education to Trade" *Prospects*, 38 (4), 449-466.

Triplett, Jack E., and Barry P. Bosworth (2006), "Baumol's disease has been cured: IT and multifactor productivity in U.S. service industries." In *The new economy and beyond: Past, present, and future,* ed. Dennis W. Jansen, 34–71. Cheltenham, UK: Edward Elgar.

Webber, D.A., and Ehrenberg, R.G. (2010). *Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education?* Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers/129/

Wellman J. (2008), *The Growing Imbalance*: *Recent Trends in U.S Postsecondary Education Finance,* Washington DC: Delta Cost Project.

Wexler, E. (2015). "Can Data Measure Faculty Productivity? Rutgers Professors Say No," *Chronicle of Higher Education*, December. HTTP://CHRONICLE.COM/ARTICLE/CAN-DATA-MEASURE-FACULTY/234595/

D.L. Wright, W.F. Fox, M.N. Murray, C.K. Carruthers, and G. Thrall (2012), "College participation, persistence, graduation and labor market outcomes: An Input-Adjusted Framework for Assessing the Effectiveness of Tennessee's Higher Education Institutions," Prepared for HCM Strategists and the Gates Foundation.