Abstract

William Massy is an emeritus professor and former Vice President for Business and Finance at Stanford University, Teresa Sullivan is President of the University of Virginia, and Christopher Mackie is a Study Director with the National Academies' Committee on National Statistics. This article summarizes the authors' recent National Research Council report, *Improving Measurement of Productivity in Higher Education,* which reviews the principles and pitfalls of measuring university productivity and proposes a practical method for doing so at the sector and institutional segment levels. The summary emphasizes the method's data requirements and describes needed changes in IPEDS and other databases.

AUTHORS

William F. Massy, Ph.D.
Stanford University

Teresa A. Sullivan, Ph.D.
University of Virginia

Christopher D. Mackie, Ph.D.
National Academies' Committee on National Statistics

CORRESPONDENCE

Email
wfmassy@comcast.net

Data Needed for Improving Productivity Measurement in Higher Education

Recognizing that higher education is a critical element of the American economy, The National Research Council of the National Academies, with support from the Lumina Foundation, convened a panel on measuring higher education productivity (NRC, 2012a). The panel members are listed in the sidebar (see next page). All are recognized experts in higher education and/or productivity analysis. This paper provides a brief summary of the panel's conclusions and, particularly, their implications for the Integrated Postsecondary Education Data System (IPEDS).

The panel's charge was to develop a practical approach for developing aggregate measures to track productivity for broad groups of institutions and for the sector as a whole. We were not asked to address the improvement of productivity itself and, likewise, research and public service productivity were outside our purview. We have proposed a conceptual structure for higher education productivity measurement. We also have documented the many difficulties and caveats associated with the use of available measurement tools.

Four touchstones guided the panel's thinking about the importance of productivity improvement and measurement:

- "Productivity should be a central part of the higher education conversation.
- Conversations about the sector's performance will lack coherence without a well-vetted and agreed-upon set of metrics.
- Quality should always be a core part of the productivity conversations, even if it cannot be fully captured by the metrics.
- The inevitable presence of difficult-to-quantify elements in a measure should not become an excuse to ignore these elements" (NRC, 2012a, p. 2).

PANEL ON MEASURING PRODUCTIVITY

Teresa A. Sullivan (Chair)
Office of the President,
University of Virginia

Thomas R. Bailey
Institute on Education and the Economy and Community College Research Center, Teachers College,
Columbia University

Barry P. Bosworth
Economic Studies Program,
The Brookings Institution,
Washington, DC

David W. Breneman
Curry School of Education,
University of Virginia

Ronald G. Ehrenberg
Cornell Higher Education Research Institute (CHERI),
Cornell University

Peter T. Ewell
National Center for Higher Education Management Systems, Boulder, CO

Irwin Feller
Department of Economics (Emeritus),
Penn State University

Barbara Fraumeni
Muskie School of Public Service,
University of Southern Maine

Juliet V. Garcia
Office of the President,
University of Texas at Brownsville and Texas Southmost College

Michael Hout
Department of Sociology,
University of California, Berkeley

Nate Johnson
Hcm Strategists,
Washington, DC

George D. Kuh
Center for Postsecondary Research (Emeritus),
Indiana University

William F. Massy
Professor Emeritus and Former Vice President for Business and Finance,
Stanford University

Carol A. Twigg
The National Center for Academic Transformation

David J. Zimmerman
Department of Economics,
Williams College

Christopher D. Mackie
Study Director

An additional requirement was that our proposed productivity measures be derived as much as possible from the discipline of economics. Productivity is defined as the ratio of outputs to the inputs required for producing them, where both inputs and outputs are adjusted for quality differences. The panel felt strongly that ad hoc measures not related to economic science, such as graduation rates and time-to or cost-of degree statistics, are incomplete and likely to be misleading when used in isolation.

Key Issues and Their Resolution

The economic definition of productivity is, fundamentally, the relation between the physical quantities of outputs and the physical quantities of inputs. It is more an engineering concept than a financial one. Financial concepts, which involve prices as well as quantities, do not enter the picture except as weights in the aggregation of disparate quantity variables (an example will be given later). The distinction matters because policies aimed at productivity improvement must address what essentially are engineering issues, which often are lost in people's concerns about financial matters. For example, a higher cost per degree that is caused by escalating labor prices (wages and benefits) does not imply reduced productivity, whereas an increase in the *amount* of labor utilized would do so.

Measures that describe either inputs or outputs, but not both, give an incomplete picture of productivity. This means that familiar statistics such as numbers of enrollments or credit hours, degree production, and cost or "profit" per faculty member should not be used to assess productivity. A final consideration is whether to evaluate "single-factor productivity" (e.g., output per labor hour) or "multifactor productivity" (output related to total resource usage). Colleges and universities have historically been labor intensive, but this has changed in recent years because of information technology, increasingly complex facilities, and outsourcing of support services (for which the labor component, while potentially significant, does not show up on the university's books). Therefore, the panel chose to measure multifactor productivity as defined by the model described later.

Educational quality is the "elephant in the room" in most discussions of higher education productivity. The economic theory is clear: both input and output quantities should be adjusted for variations in quality. For example, because of changes in technology, today's cars and computers are not directly comparable to those produced a decade ago; thus, direct price comparisons (without quality adjustment) are not meaningful. Therefore, when measuring price and productivity changes in these sectors, economists use techniques that account for changes in input and output characteristics from one period to the next.

Productivity measurement outside higher education, in competitive markets, relies on one of two devices to police quality. First, other things equal, better quality usually commands higher prices (the more expensive car or computer is usually the better one). It turns out, however, that this approach does not work for higher education. The prevalence of government subsidies and regulation, coupled with a dearth of well defined and accurate market information about education quality (particularly as it pertains to learning), make it unwise to assume that either the tuition rates or financial aid awards of colleges and universities are determined by competitive market forces.

The other method is to measure quality through special studies – for example, to track differences in the speed and memory capacity of computers – and use the resulting measures to adjust the quantity variables. The panel looked carefully at the prospects for developing the kinds of comprehensive learning quality measures needed to make such adjustments. We would have liked nothing better than to propose such measures but, unfortunately, we were forced to conclude that this will not be possible anytime soon. Our report cites a great deal of good work in the area, which definitely should be continued; but, while current and prospective learning and engagement measures are useful in particular contexts, they cannot be brought together into comprehensive, robust, indices for quality adjustment.

All is not lost, however. In the United States, a variety of external quality assurance procedures are deployed, such as regional accreditation, subject-specific accreditation, and in some fields, stringent licensure requirements. To the extent these work as designed (and they should be made to work regardless of whether productivity is measured or not), they put a floor under the output quality of education. Institutions also employ a variety of internal quality assurance procedures based, for example, on faculty governance. When combined with the external procedures, these can be expected to deter any "race to the bottom" that might result from measuring the quantitative aspects of higher education production. While it is true that high-end quality differences will not be reflected in quantity-based productivity statistics, at least the downside dangers can be mitigated. We hope that, in due course, better and more comprehensive quality measures will shed light on how learning varies across segments and changes over time. In the meantime, subjective judgments can be used to interpret the quantitative productivity statistics in light of more fragmented evidence about output quality.

The panel focused on instruction rather than on research and public service—even though the latter are central to the mission of a large subset of institutions. Including the research mission would have carried us into territory already being considered by another National Academies panel (NRC, 2012b) and, in any case, it would have added huge complexities to the ones already confronting us. But, while the panel did not address research and public service productivity, we did carefully consider their impact on the measurement of educational productivity. Inputs must be parsed into their instruction and research/public service components. The parsing is mostly a straightforward application of cost accounting but, as mentioned later, we do propose a new approach for handling the vexing issue of departmental research.

The economic definition of productivity is, fundamentally, the relation between the physical quantities of outputs and the physical quantities of inputs. It is more an engineering concept than a financial one… The distinction matters because policies aimed at productivity improvement must address what essentially are engineering issues, which often are lost in people's concerns about financial matters.

Another issue is the importance of avoiding spurious comparisons among institutions with dramatically different characteristics and missions. Among other things, it is essential to take into account incoming student ability and preparation. For example, highly selective institutions typically have higher completion rates than open-access institutions. This may reflect more on the prior learning, preparation, and motivation of the entrants than on the productivity of the institutions they enter—which means institutional performance should be gauged in terms of value added, not the absolute quality of graduates. Therefore, for the purpose of generating performance statistics, institutions should be segmented into reasonably homogeneous categories – for example, as used in the Carnegie classification system and the Delta Cost Project (2009).

The list of measurement issues for the sector would not be complete without consideration of data availability. The panel adopted a two-pronged approach. Our "Base Model," described in the next section, relies almost entirely on current IPEDS and other government datasets. The "enhanced model" that follows requires additions to IPEDS. As explained later, we believe these additions will be worthwhile for their own sake as well as to improve productivity measurement. The paper ends with a brief description of the other data-related recommendations in the panel's report.

Base Productivity Measurement Model

The panel's conceptual framework employs a multifactor productivity index, the so-called "Törnqvist index." The Bureau of Labor Statistics (BLS) and the Organization for Economic Cooperation and Development (OECD) use this index to measure productivity in a variety of economic sectors (BLS 2007; OECD 2001, which includes references to the background literature). We first describe the index in general terms, then define the output and input variables, and, finally, illustrate the calculations with a numerical example.

The productivity index, as evaluated for time increment Δt, is:

$$\textit{Productivity index}\ [\Delta t] = \textit{Output index}\ [\Delta t] \div \textit{Input index}\ [\Delta t].$$

The input and output indices represent changes in the physical quantities over the time increment Δt (e.g., from 2010 to 2011). In other words, the Törnqvist index defines *Productivity* as the change in outputs obtainable from the input changes observed over Δt. *Productivity change*, in turn is calculated from the ratio of successive productivity indices:

$$\textit{Productivity change } [\Delta t_1 \text{ to } \Delta t_2] = \textit{Productivity index } [\Delta t_2] \div \textit{Productivity index } [\Delta t_1] - 1.$$

Colleges and universities have historically been labor intensive but this has changed in recent years because of information technology, increasingly complex facilities, and outsourcing of support services.

These definitions are consistent with the conceptualization of productivity as an engineering concept. Productivity is the slope of the "production function"—the curve relating outputs to inputs. Looking at the slope of the function rather than the function itself amounts to a kind of "what if" analysis: what happens to outputs if the inputs change by a certain amount?

Outputs. We recommend a simple yet comprehensive measure for output quantity. It is based on two IPEDS variables that, in the panel's words, "are the standard unit measures of instruction in American higher education."

- *Credit hours*: 12-month instructional activity credit hours summed over student levels (e.g., undergraduates, first professional students, and graduate students);

- *Completions:* awards or degrees conferred, summed over programs, student levels, race or ethnicity, and gender (NRC, 2012a, p. 65).

The importance of completions is obvious, but a measure based only on completions would ignore the learning that takes place on a course-by-course basis. The panel's recommendation, therefore, is that the base definition of educational output be "Adjusted credit hours" (ACH), defined as follows:

$$\textit{Adjusted credit hours} = \textit{Credit hours} + (\textit{Sheepskin effect} \times \textit{Completions}).$$

Again to quote the panel, "The 'sheepskin effect' represents the additional value that credit hours have when they are accumulated and organized into a completed degree. Based on studies of the effect of earned credits and degrees on salaries, the panel believes that a value equal to a year's worth of credits is a reasonable figure to use as a placeholder for undergraduate degrees. Additional research will be needed to determine the appropriate weight for the sheepskin effect for graduate and 1st professional programs" (NRC, 2012a, p. 66). The same ideas apply to many community college programs.

Inputs. The model's inputs, defined in Table 1, are the quantity of labor, the amount of non-labor expenditure, and the rental value of capital (the depreciation of plant and equipment during use) utilized in the educational process. The data for each input consists of (a) physical quantities or surrogates for quantities, as required by the fundamental definition of productivity; and (b) nominal expenditures, which are used to combine the several inputs into a single index.

Labor FTEs is a direct physical quantity. No such physical quantity can be found for *Expenditures on intermediate inputs* or *Capital*, so for them it is necessary to use deflated dollars as surrogates. (The dollar figures sum up the myriad of individual items included in the definition for each variable, weighted by the items' unit costs.) To summarize, the model uses three input variables (labor, intermediate inputs, and capital), each of which is represented by a physical quantity (or surrogate) and nominal expenditures.

As stated earlier, the input variables must reflect instruction rather than the whole range of institutional activity. The Delta Cost Project (2009) and OMB Circular A21 provide the needed methodology. The steps, which must be performed for each variable in Table 1, are to: (1) separate total expenditures into their direct and indirect (administrative and support) components; (2) further separate the direct component into instruction, research, and public service; (3) allocate the indirect component according to the fractions of direct expenditures; and (4) add the indirect allocation for instruction to the direct instruction variable to produce an overall figure for instruction.

Table 1

Input Variables Used in the Proposed Model

a) *Expenditures on Labor* (LE): nominal value of salaries and wages plus fringe benefits, used as the weight of L when aggregating the input.
b) *Labor* (L): the quantity measure for labor input, approximated by full-time equivalent (FTE) employees. Both academic and non-academic employees are included in the calculation. FTE figures are calculated from total full- and part-time employees, with a part-time employee counting as one third of a full-time employee, as assigned in IPEDS (this, too, could be adjusted with empirical justification). Labor is the biggest input into higher education instruction.
c) *Expenditures on Intermediate Inputs* (IE): Nominal cost of materials and other inputs acquired through purchasing, outsourcing, etc. (the sum of the IPEDS 'operations & maintenance' (O&M) and 'all other' categories). These nominal values are used in calculating I weights.
d) *Intermediate Inputs* (I): Deflated nominal expenditures (IE) are used to represent the physical quantities.
e) *Expenditures on Capital* (KE): opportunity cost for the use of physical capital; also called rental value of capital. Expenditures equal the IPEDS book value of capital stock times an estimated national rate of return on assets, where book value capital stock equals the sum of land, buildings, and equipment. Overall, the book value reported in IPEDS is likely too low; however, it does include buildings that may not be specifically allocated to teaching, which offsets the total to an unknown degree. These nominal capital values are used in calculating the weighting for capital.
f) *Capital* (K): For the quantity of capital input, the book value is deflated by the Bureau of Economic Analysis's investment deflator for gross private domestic investment.

The deflators for intermediate expenditures and capital are, respectively, the Producer Price Index and the index for gross private domestic investment. Neither these figures nor the national rate of return on assets can be obtained from IPEDS, but they are available from other government sources.

Calculating the index. Table 2 illustrates the index's calculation. The table reflects three years of data for one institution, which allows representation of two Δt increments. The time periods need not reflect a single year and the intervals between periods need not be the same for all Δt increments. (Scaling adjustments are needed if the increments vary, however.) Given the period-to-period variability inherent in IPEDS data, for example, it may be desirable to consider each period as the average of, say, three to five years. We noted earlier that the productivity index should not be applied to single institutions, but rather to broad groupings of institutions (the "segments" discussed above). The availability of institution-specific IPEDS data makes this relatively easy, and the aggregation process further mitigates the data variability problem.

Other things equal, better quality usually commands higher prices (the more expensive car or computer is usually the better one). It turns out, however, that this approach does not work for higher education.

The output calculations appear at the top of the table. Credit hours and completions are extracted directly from IPEDS. Adjusted credit hours (ACH) are calculated using a sheepskin effect of 30 (one year's normal student load). Quantity change is obtained by dividing each ACH figure by the preceding one: e.g., 1.008 = 643,477 ÷ 638,435. This also equals the output index because enrollments and completions already have been combined into a single number by applying the sheepskin effect.

Table 2

Base Model Productivity Calculations

	Period 1	Period 2	Period 3
Output			
Credit hours	573,815	574,176	602,000
Completions	2,154	2,310	2,500
Adjusted credit hours (ACH)	638,435	643,476	677,000
Quantity change (= output index)		1.008	1.052
Inputs			
Quantities			
Labor FTEs (L)	3,926	4,296	4,705
Intermediate expenditures (I)	$324,680	$486,147	$643,599
Rental value of capital (K)	$261,834	$267,507	$348,033
Expenditures			
Wages and fringe benefits (LE)	$756,399	$867,311	$1,036,594
Intermediate expenditures (EI)	$324,680	$550,921	$777,193
Rental value of capital (KE)	$261,834	$301,954	$400,791
Total	$1,342,913	$1,720,186	$2,214,578
Quantity change			
Wages and fringe benefits		1.094	1.095
Intermediate expenditures		1.497	1.324
Rental value of capital		1.022	1.301
Expenditure weightings			
Labor FTEs		53.4%	48.6%
Intermediate expenditures		28.1%	33.6%
Rental value of capital		18.5%	17.8%
Total		100.0%	100.0%
Input index		1.180	1.204
Multifactor productivity			
Productivity index		0.854	0.874
Change in productivity		2.3%	

The "sheepskin effect" represents the additional value the credit hours have when they are accumulated and organized into a completed degree.

The situation is more complicated for inputs. First, as described earlier, the IPEDS quantity and expenditure variables must be adjusted to account for research and public service. The resulting quantity change figures are ratios of the successive quantity figures (e.g., 1.094 = 486,147 ÷ 324,680), but then they must be combined to produce a composite input index. This is accomplished using a weighted geometric average with the relative expenditure figures (also shown in the table) as weights. For example, 1.180 is the geometric average of 1.094, 1.497, and 1.022. (The weights for Intermediate expenditures and Capital in Period 1 are the same as the quantities, but they differ in later years because of the inflation adjustment.) The choice of a geometric average follows from the fact that ratios are being averaged, and also from the mathematical derivation of the Törnqvist index.

Calculation of the multifactor productivity index appears at the bottom of the table. The index itself is the ratio of the output index to the input index: e.g., 1.008 ÷1.180. The change in productivity is the ratio of the successive productivity indices minus one: in this case the shift from 0.847 to 0.854 is a 2.3% change.

The Törnqvist index has some very desirable properties. In particular, researchers have shown that, under fairly general conditions, this calculation makes the best possible use (in terms of productivity measurement) of the information embedded in the input and output variables. Among other things, it washes out extraneous financial factors like the effects of substituting one resource for another because of price changes. (Substituting computers for people in a production process simply because the former have become relatively cheaper does not represent a productivity increase, for example, whereas making the substitution because the computers are being used more effectively does.) Alternative methods of calculation—for example, weighting the input changes by something other than nominal expenditures or using an arithmetic average instead of a geometric one—have been shown to produce inferior results. The panel recommended that a task force be set up to work on operationalizing our conceptual structure, and that it begin with consideration of the Törnqvist index.

The panel's charge called for addressing productivity measures at different levels of aggregation including the institution, system, and sector levels. Our proposed model is designed to operate at the sector or subsector (segment) level. While we feel obligated to raise the possibility of single-institution and state-specific indices (which are possible given that IPEDS provides institution-specific data), we do not want to invite use of the model for accountability purposes. To do so would produce malevolent incentives–and, possibly, a race to the bottom in terms of education quality. Moreover, state-level aggregations would necessarily obliterate mission-distinctions that are typically important within each state.

Enhancements to the Base Model

While the panel believes the base model to be a viable approach for tracking college and university productivity, we have identified two enhancements that, while requiring data collection beyond the current IPEDS structure, would add significantly to the model's power.

We do not want to invite use of the model for accountability purposes. To do so would produce malevolent incentives – and, possibly, a race to the bottom in terms of education quality.

Differentiating labor categories. The first enhancement is to track key labor categories, in our case academic versus non-academic and regular versus casual employees, separately from total FTEs. The panel's reasons for suggesting this differentiation are as follows (see NRC, 2012a, p. 74):

1. One of the critical assumptions of the conventional productivity model is not viable in higher education. The typical productivity study assumes that, because labor is procured in competitive markets, relative compensation approximates relative marginal products. There is, in such a situation, no need to differentiate labor categories. Unfortunately, tenure-track faculty labor may not be linked tightly to marginal product in education because such faculty often are valued for research and reputational reasons, or be protected by or locked into institutions by tenure. Similarly, many so-called "contingent" faculty (hired on a course-by-course basis, often without fringe benefits) are desired as an alternative to tenure rather than because of judgments about their marginal product.

2. Another assumption is that the market effectively polices output quality. This is manifestly not the case for higher education. Colleges pursue strategies—larger classes or less costly instructors, for example—that reduce cost per nominal output but which will dilute quality when taken to extremes. In the example above, it may be attractive to employ less expensive, but also less qualified, personnel who are not well integrated into institutional quality processes. The panel is concerned lest the measurement of productivity add to the already problematic incentives to emphasize quantity over quality in higher education.

3. Academic staff play a unique and critically important role in most institutions. They, and only they, can make the intellectual judgments needed to create new knowledge, decide curricular and pedagogical issues, and assure educational quality. It is true that the distinction between teaching and nonteaching staff blurs as information technology shifts the modalities of teaching and learning. In some institutions, for example, faculty time is leveraged by modern learning software, a change that may require entirely new kinds of labor inputs. Such technology-driven changes are not unique to higher education, but the pace of change seems unusually brisk at the present time. Yet the critical task of intellectual leadership remains with faculty. Singling them out as a separate labor category recognizes that, at root, this kind of "labor" is not truly substitutable.

4. Productivity statistics are more likely to weigh heavily in policy debates on higher education than in policy debates for other industries. The U. S. public policy environment includes a significant oversight and accountability component that

requires information about productivity. Therefore, it is important that the statistics be as complete as possible on the important issues, including those associated with labor substitution.

Differentiating labor categories will require IPEDS be modified to include the following three-way classification scheme.

1. *Regular faculty FTEs*: those on the tenure line or equivalent, whether full or part-time.

2. *Part-time teachers who are hired on a course-by-course basis:* may be measured in terms of FTEs, number of course assignments, or some other metric–perhaps related to the current "Part-time" and "Primarily-instruction" ("PT/PI") variable.

3. *All other FTEs*: total FTEs excluding the above.

Importantly, the new scheme will need to support allocations among education, research, and public service as described earlier.

The panel was able to approximate the three-way differentiation from current IPEDS data (see NRC, 2012a, p. 77), but we believe it will be worthwhile to produce the data directly. (Development of the methodology should involve consultation with the providers and users of IPEDS data.) In addition to their value for productivity measurement, these data will prove useful to developers of benchmarking statistics and to others in the higher education research community.

The U.S. public policy environment includes a significant oversight and accountability component that requires information about productivity.

Differentiating outputs. Controlling for output heterogeneity is the other important enhancement. The resources required to produce an undergraduate degree vary significantly across fields, and the cost of bachelors' degrees differs systematically from the costs of associate, graduate, and first professional degrees. Failure to control for these differences would risk mistaking output shifts from the more expensive disciplines of science, technology, engineering and mathematics (STEM) to non-STEM disciplines for a decline in productivity.

Enhancing the model in this way will require the data for both degrees and enrollments to be differentiated by field and award level. This poses no problem because IPEDS already provides the needed data. Differentiating enrollments is a different story, however, because IPEDS does not report credit hours by field. Institutions may track credit hours by the department or discipline in which the course is taught—in order to apply the Delaware cost benchmarks, for example—but these data cannot be mapped directly to degree production because students take many courses outside their matriculated areas. Researchers have made the necessary correspondences by creating course-taking profiles for particular degrees, but these matrices are difficult to manage and maintain on an institution-wide basis.

A better way is to collect data in a way that follows the students, not just the departments that teach them. The information needed to do this exists in most institutions' student registration files. Extraction of the needed data could proceed as follows:

- "Identify the students matriculated in a given degree program ('output category') as defined by the IPEDS fields and degree levels. Undeclared students and students not matriculated for a degree would be placed in separate ('non-attributable') output categories.
- For each output category, accumulate the credit hours earned by the students in that category, regardless of the department in which the course was offered or the year in which it was taken.
- Allocate credits earned by matriculated but undeclared students in proportion to the credit-hour fractions of declared students for the given degree. Retain non-matriculated students in their own separate category, one that has no sheepskin effect but in other respects is treated the same as the other categories" (NRC, 2012a, p. 78).

Once again, the value of such statistics will greatly transcend productivity measurement. One can envision new approaches to the analysis of graduation rates and times to degree, for example, and credit hour measures that "follow the student" will permit more accurate costing measures to be produced within institutions. As noted above, we believe the requisite data exist within existing institutional data files. Hence all that is necessary is to develop appropriate extraction algorithms.

Other Data-Related Recommendations

The NRC panel's report offers a number of additional recommendations pertaining to the definition and development of datasets. The importance of most of these is self-evident, and readers should refer to the original report for additional discussion.

Academic staff play a unique and critically important role in most institutions. They, and only they, can make the intellectual judgments needed to create new knowledge, decide curricular and pedagogical issues, and assure educational quality.

The first such recommendation states that, "Definitions should be established for outcomes and institutions other than traditional four-year colleges and universities with low transfer-out rates, and appropriate bonus figures estimated and assigned to those outcomes. This is especially important for community colleges where, in contrast to BA and BS degrees, outcomes may be successful transfers to 4-year colleges, completion of certificates, or acquisition of specific skills by students with no intention of pursuing a degree" (NRC, 2012a, p. 92).

Two recommendations address the handling of research in universities where it is a major mission element. First, "NCES or a designee should develop an algorithm for adjusting labor and other inputs to account for joint production of research and service. Faculty labor hours associated with instruction should exclude hours spent on sponsored research and public service, for example and the algorithm should provide an operational basis for adjusting other inputs on the basis of expenditures" (NRC, 2012a, p. 95). Our conceptual framework already provides an algorithm for separately budgeted research and service activities (based on the Delta Cost Project and A21), but it may be possible to refine the approach.

The treatment of "departmental" (i.e., not separately budgeted, including sponsored) research is considerably more difficult. Such research is paid for by the university, often in the form of teaching-load reductions. The panel recommends the development of a statistical model to parse departmental research (DR) into two components: *Project-driven* and *Discretionary*. Such a study would use sample data on faculty activity to build statistical models for estimating: (a) the amount of activity that should be classified as departmental research, by institutional type, field, amount of sponsored research, and other descriptors; (b) the share of that activity directly associated with sponsored projects; and (c) the (residual) share of activity that should be classed as discretionary. We believe such a study will turn out to be practical. It would be immensely valuable for costing and other purposes as well as for productivity measurement.

As stated in the report, "The direct link between *Project-driven DR* and sponsored research provides a strong argument for excluding the former from instructional costs. Only the idiosyncrasies of university accounting and the market power of sponsoring agencies enable those agencies to enforce cost-sharing on academic-year effort in order to spread their funds further. Arguing on principle for inclusion of research cost and instructional cost is tantamount to arguing that the sponsored research itself should be included–which, in addition to being intrinsically illogical, would hugely distort the productivity measures" (NRC, 2012a, p. 97).

Discretionary DR, on the other hand, refers to work initiated by faculty members without regard to external support. In the panel's view, "Good arguments exist for including at least a part of such activity in the cost base for instruction. For one thing, it is difficult or impossible to separate the effects of educational research and development (R & D) from the other motivators of low teaching loads (other than those associated with sponsored research projects), and there is no doubt that educational R & D should be included in the instructional cost base. Meaningful education R & D expenses and work that sustains the life of disciplines (and is not sponsored research) should be defendable to stakeholders.

A better way is to collect data in a way that follows the students, not just the departments that teach them. The information needed to do this exists in most institutions' student registration files...Hence all that is necessary is to develop appropriate extraction algorithms.

Additionally, some allocation of faculty time entails service that is required to keep institutions running" (NRC, 2012a, p. 99).

Three more recommendations address more general data questions pertaining to higher education. The first says that, "Every effort should be made to include colleges and universities in the U. S. Economic Census, with due regard for the adequacy of alternative data sources and for the overall value and costs added, as well as difficulties in implementation" (NRC, 2012a, p. 110). Colleges and universities were included in the economic census only once, in 1977. Considering the importance of the sector, the amount of resources it consumes, and recent advances in university data systems, we could find no reason for continuing the exemption.

Another recommendation states that, "Standardization and coordination of states' student record databases should be a priority. Ideally, NCES should revive its proposal to organize a national unit record database" (NRC, 2012a, p. 117). We recognize that this is a politically difficult recommendation that may take years to realize, but emphasize that such a system, with appropriate privacy safeguards, would be extremely valuable for both policy and research purposes.

Finally, the panel supported the idea that, "The Bureau of Labor Statistics should continue its efforts to establish a national entity such as a clearinghouse to facilitate multi-state links of unemployment insurance (UI) records and education data. This would allow for research on issues such as return on investment from postsecondary training or placement rates in various occupations" (NRC, 2012a, p. 118). The importance of state longitudinal databases, the national student clearinghouse, and various survey-based databases also are discussed in the report.

In closing, we hope that concerns about productivity measurement in higher education reflected in the charge to the panel continue to receive the attention they deserve. For example, a government entity such as the Department of Education, Census Bureau, or Bureau of Labor Statistics should be charged with overseeing and testing the implementation of our conceptual framework. We hope that such a group will be formed without delay, and also that individual institutions will begin to experiment with some of the data enhancements described in this paper.

References

Bureau of Labor Statistics (BLS). (2007). *Technical information about the BLS multifactor productivity measures*. Available at http://www.bls.gov/mfp/mprtech.pdf

Delta Cost Project. (2009). *Metrics for improving cost accountability* (Issue Brief No. 2). Washington, DC: Author. Available at http://www.deltacostproject.org/resources/pdf/issuebrief_02.pdf

Massy, W. F., Sullivan, T. A., & Mackie, C. (2013). Improving measurement of productivity in higher education. *Change*, January/February.

National Research Council (NRC). (2012a). *Improving measurement of productivity in higher education*. Panel on Measuring Higher Education Productivity: Conceptual Framework and Data Needs. T. A. Sullivan, C. Mackie, W. F. Massy, & E. Sinha (Eds.). Committee on National Statistics and Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

National Research Council (NRC). (2012b). *Research universities and the future of America: Ten breakthrough actions vital to our nation's prosperity and security*. Committee on Research Universities; Board on Higher Education and Workforce; Policy and Global Affairs; National Research Council. Washington, DC: The National Academies Press, June.

Organization for Economic and Cooperative Development (OECD). (2001). *Measuring productivity, measurement of aggregate and industry-level productivity growth*. Paris: Author.