This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Productivity in Higher Education

Volume Authors/Editors: Caroline M. Hoxby and Kevin Stange, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-57458-5 (cloth); 978-0-226-57461-5 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/productivity-higher-education

Conference Date: May 31–June 1, 2016

Publication Date: November 2019

Chapter Title: Introduction to "Productivity in Higher Education"

Chapter Author(s): Caroline M. Hoxby, Kevin Stange

Chapter URL: https://www.nber.org/books-and-chapters/productivity-higher-education/introduction-productivity-higher-education

Chapter pages in book: (p. 1 – 15)

Introduction

Caroline M. Hoxby and Kevin Stange

Our Enterprise

One of us (Hoxby) recalls a meeting, not so long ago, in which university leaders and faculty were discussing a vast project that cost at least nine figures. The costs were discussed in great detail. As the discussion neared its end, Hoxby finally asked, "But what are the benefits of the project? Is its ratio of benefits to costs high, or would it be better to allocate the funds to more productive uses?" These questions startled the group for two reasons. First, those assembled had fallen into the habit of associating the merit of a project with its costs, not its ratio of benefits to costs (its productivity). Second, most thought it absurd even to consider measuring benefits. These two reasons were related: because most believed that it was impossible to measure benefits, they routinely focused on costs. Indeed, these habits are not limited to university staff. When higher education experts were asked which was the best institution, they tended to suggest a costly one and cite its high spending as evidence of its quality.

To economists at least, it seems unnatural to think so much about costs but so little about the productivity of a sector, higher education, that plays such a crucial role in the economy and society. For any society-wide question that involves allocating resources between some other sector, such as health

Caroline M. Hoxby is the Scott and Donya Bommer Professor in Economics at Stanford University, a senior fellow of the Hoover Institution and the Stanford Institute for Economic Policy Research, and a research associate and director of the Economics of Education Program at the National Bureau of Economic Research.

Kevin Stange is associate professor of public policy at the University of Michigan and a research associate of the National Bureau of Economic Research.

For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see http://www.nber.org/chapters/c13874.ack.

You are reading copyrighted material published by University of Chicago Press. Unauthorized posting, copying, or distributing of this work except as permitted under U.S. copyright law is illegal and injures the author and publisher.

care, and higher education, we need to know the sectors' relative productivity. When judging whether the market for higher education generates good incentives or, rather, is plagued by market failures that allow institutions to be grossly inefficient, we need to know productivity. When assessing government policies, such as grants or loans, that subsidize students, we need to know the productivity of the investments these policies facilitate. To allocate a budget efficiently among their institution's many activities, higher education leaders need to understand productivity. When students decide whether and where to attend college, they need to know whether those investments will be productive. Thus, at every level of decision-making (social, institutional, individual), the productivity of higher education investments is crucial.

This volume, *Productivity in Higher Education*, is the result of a concerted effort by National Bureau of Economic Research scholars to advance the frontier of knowledge about productivity in higher education. The timing of this push is not accidental. Rather, it is the result of newly available data that allow us to assess benefits *much* better and analyze costs better as well. The new data come from administrative sources and therefore tend to be accurate. They also tend to be population data, not a sample. These attributes of the data are crucial for many of the studies in this volume. It is not merely that the better data make the findings more precise or permit otherwise infeasible empirical strategies, though they do both of these. Crucially, they allow researchers to ask questions that simply could not have been asked previously. This expansion of the frontier of questions we can credibly answer meets a heightened demand for these answers from students, parents, and policy makers. We think it is fair to say that the productivity of higher education institutions—from large elite research universities to small for-profit colleges—has never been under greater scrutiny than right now. In short, this is an exciting and opportune time for research on productivity in higher education.

This is not to say that the enterprise is without challenges. Some of these challenges recur so often among the studies in this volume that it is worthwhile enumerating them now. The first challenge is multiple outcomes. Higher education potentially affects skills, earnings, invention, altruism, employment, occupations, marriage, and many other outcomes. Even if we have data on all such outcomes, how are researchers to prioritize them for analysis? A related difficulty is the "multiproduct" nature of institutions of higher education. Even the simplest institutions usually have several degree programs, and large research universities conduct a bewildering array of activities across numerous major domains: undergraduate teaching, professional programs, doctoral advising, research, medical experimentation, and so on. Any study of higher education must prioritize which activities to analyze and make thoughtful decisions about how to allocate costs associated with resources, such as infrastructure, that are shared by several activities.

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

The third recurrent challenge, which can be formidable, is selection on students' aptitude and prior achievement. By the time they arrive at the door of a postsecondary institution, students are already greatly differentiated. Indeed, students who attend the most-selective universities have arguably learned as much by the end of the eighth grade as have the high school graduates who attend the least selective institutions. Any study that credits the most-selective institutions with the incoming preparation of its students will greatly exaggerate the productivity of those institutions. Similarly, a study might exaggerate the productivity of an institution if it draws students from richer backgrounds and these backgrounds have an independent effect on future outcomes. A fourth recurrent challenge is attribution. Students who attend nonselective institutions often initially enroll at one school, take some classes at another, and finally finish at a third. Students who attend highly selective colleges do not "roam" so much as undergraduates, but they often pursue graduate education. All these layers of education affect a person's outcomes, making attribution to any one institution difficult. A fifth recurrent challenge is the public nature of some potential benefits of higher education. Public benefits range from students learning from one another (at the micro end) to research findings that benefit all mankind (at the macro end). While these benefits may be measurable, it is often difficult to trace them to their source. A final recurrent issue is that there are always at least a few perspectives from which to assess productivity. There is the view of society as a whole, personified as the social planner who takes into account all the benefits of higher education, including public ones, but also takes into account all the costs, regardless of who pays for them or how. Then there is the view of a government: Will its current spending on higher education generate sufficient future tax revenue to balance the books? Finally, there is the view of students. They may care little about public benefits and will surely focus on costs they pay themselves versus those funded by taxpayers or philanthropists.

In common, the studies in this volume confront these challenges. Each study deals with them differently, and that is part of what makes the collection interesting. Together, the studies constitute a concise course in approaches to overcoming these challenges. But is there anything that these approaches share?

First, all the studies in this volume are resolutely practical when approaching the challenges mentioned above. The authors refuse to give up on measuring benefits or costs simply because the measures are imperfect. They make smart choices or, when the choice is not obvious, adopt a pluralistic attitude and offer several reasonable measures as alternatives. When the authors choose which outcomes to prioritize, they pay attention to those that appear to matter most to those who finance higher education. The authors are also practical in consistently focusing on institutions' "core business," the activities that are most important to the schools they are studying. (These

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

are not necessarily the same activities for, say, a for-profit online institution versus a selective research university.)

Second, the studies in this volume are scrupulous about differentiating between evidence that is credibly causal and evidence that is only correlational. Each study devotes great effort to developing an empirical strategy that can produce results that can confidently be identified as causal. Some of the studies use the equivalent of randomized controlled trials. Others rely on natural experiments. Others must rely on nonexperimental and descriptive analysis. Regardless, all the studies are frank about which results can confidently be given a causal interpretation and which should be read more cautiously.

Finally and most importantly, all the studies draw deeply on economic reasoning. Prior work on productivity in higher education has not used economics to structure hypotheses and evidence. But at least in the United States, economics does apply to the higher education sector. Institutions do function in markets. They encounter prices set by others and set prices themselves. They face incentives, and their workers act in accord with incentives. Institutions cannot set faculty salaries or tuition arbitrarily but are constrained by supply and demand. Their nontuition sources of revenue (grants, gifts, appropriations, patent revenues) also depend, in a fairly systematic way, on their producing certain outcomes. Students may be less informed investors than is optimal, but they do make choices among institutions and decide whether to stay enrolled or leave. In short, higher education has its own labor markets, its own investors, and its own industrial organization. By drawing systematically on insights from labor economics, finance, and industrial organization, researchers make much more progress than would otherwise seem possible. (Admittedly, the analysis would be easier if all institutions of higher education shared a straightforward objective function such as maximizing profits.)

The Challenges Are Not Unique to Higher Education

We enumerated five major challenges to understanding productivity in higher education: multiple outcomes, the multiproduct nature of institutions, selection, attribution, and the public nature of some potential benefits. In a very insightful follow-up to this introduction, in chapter 1, "What Health Care Teaches Us about Measuring Productivity in Higher Education," Douglas Staiger explains how the *same* challenges plague the study of productivity in health care. By showing us the parallels, he clarifies each challenge and allows us to see it at a 1,000-foot view, not a view too close for acumen. Moreover, he lays out how health care research has addressed each challenge, thereby giving us highly practical guidance. We encourage readers to read Staiger's contribution as a second introduction that will organize their thinking on all subsequent chapters.

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

What We Learned

In chapter 2, "The Productivity of US Postsecondary Institutions," Caroline Hoxby attempts to compute the productivity of the vast majority of undergraduate programs in the United States. In the process, she addresses several of the key issues that plague studies of productivity in higher education. For instance, she argues that the multiple outcomes problem cannot be evaded but can be addressed by presenting results based on a variety of outcomes. The study emphasizes productivity results based on earnings (because these matter disproportionately for the financial stability of the postsecondary sector) but also shows productivity results based on public service and innovative contributions.

The study's most important advance is a proposed remedy for the selection problem that is based on comparing the outcomes of students who are extremely similar on incoming achievement (test scores, grades) and who, crucially, apply to the same postsecondary institutions, thus demonstrating similar interests and motivation. This approach employs all the possible quasi experiments in which a student "flips a coin" between schools that have nearly identical selectivity or in which admission staff "flip a coin" between students with nearly identical achievement. This quasi-experimental remedy is intuitive and credible, but it also generates extraordinary data requirements. Thus this study gives us our first example of how having better data allows us to pursue empirical and econometric strategies that would otherwise be out of bounds.

The longitudinal data in this study allow Hoxby to compute *lifetime* educational costs (private and social) and predict *lifetime* benefits.

The study's most important finding is that when earnings are used to measure benefits, the productivity of a dollar is fairly similar across a wide array of selective postsecondary institutions. This result is striking because the most-selective schools spend several times as much per pupil as schools that are only modestly selective. That is, educational resources and students' capacity to use those educational resources are rising in a sufficiently parallel way that productivity is roughly flat—even though selectivity and resources both rise fairly dramatically. This result indicates that there are no easy gains to society from, say, taking a dollar away from the most-selective institutions and giving it to somewhat less-selective ones. Also, this result suggests that market forces compel some amount of efficiency among selective institutions. However, Hoxby also finds that compared to selective institutions, nonselective postsecondary institutions are less productive on average and vary greatly in their productivity. This result implies that market forces exert little discipline on such schools, allowing nonproductive institutions to attract students even when they are located side by side with much more productive ones.

Interestingly, the study also concludes that market forces do not discipline

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

postsecondary institutions if public service is used as the measure of institutional output: selective schools that enroll very similar students and cost very similar amounts differ substantially on their contributions to public service.

The multiple outcomes problem is also tackled by Veronica Minaya and Judith Scott-Clayton in chapter 3, "Labor Market Outcomes and Postsecondary Accountability: Are Imperfect Metrics Better Than None?" This chapter exemplifies the relentless practicality described as a prime virtue of studies in this volume. Minaya and Scott-Clayton put themselves in the shoes of policy makers who want to assess their state's institutions but who can access only the data that could likely be made available. They use an impressive database that contains demographics (race, ethnicity, sex, age), zip code at initial enrollment, and full postsecondary transcripts for all students who enrolled in a public institution in one of the most populous US states over an 8-year period. They then follow the students for 10 years in the state's employment, unemployment, and earnings records. They face realistic constraints, such as students being "lost" if they transfer to private colleges or move out of state.

What could a policy maker learn about institutions' productivity from such data? Minaya and Scott-Clayton's first key finding is that transcript data are insufficient. While transcript data allow them to construct productivity measures based on credits earned and degree attainment, schools' rankings change substantially when outcomes based on labor market data are added. This indicates that the skills students learn are not fully summarized by what their transcripts say, especially if we weigh skills by how employers value them. Moreover, Minaya and Scott-Clayton find that there are important improvements in knowledge about productivity if we do not merely rely on *early* labor market–based measures (earnings and employment in initial jobs) but observe the whole first decade of a student's career.

The authors' other key finding is that productivity measures are problematic if they do not adjust for students' demographics and the sociodemographics of the zip code from which they come. Unadjusted measures overstate the productivity of institutions that draw students whose incoming demographics likely give them advantages in college and in finding initial jobs. Tellingly, *adjusted* short-term measures are more correlated with long-term labor market outcomes. This is a classic test of whether a measure truly records value added or whether it reflects incoming differences (selection). The reason this test works is that incoming differences, such as whether a student grew up in a richer family, are valued less by the labor market over time, whereas skills are valued as much or more over time (Altonji and Pierret 2001).

A similar test is used by Evan Riehl, Juan E. Saavedra, and Miguel Urquiola, who analyze administrative data that includes, remarkably, learning outcomes. For chapter 4, "Learning and Earning: An Approximation to College

You are reading copyrighted material published by University of Chicago Press. Unauthorized posting, copying, or distributing of this work except as permitted under U.S. copyright law is illegal and injures the author and publisher.

Value Added in Two Dimensions," they draw on data from Colombia, a country with a vigorous market for higher education that is not dissimilar to that of the United States. Importantly, Colombian students' learning is assessed by standardized examinations not only before they enter universities but also when they exit. Although the exit examinations are partially field-specific, parts of the preentry and exit examinations are designed to measure progress on a core set of skills. Thus the exams generate natural, learning-based measures of institutional value added. As a result, Riehl, Saavedra, and Urquiola assemble a uniquely comprehensive set of outcomes: outcomes based on transcripts (which provide important outcomes such as degree completion but are hard to compare across institutions); outcomes based on earnings and employment; and outcomes based on measures of learning that are standardized across institutions. We cannot overemphasize how singular this situation is. It is not merely that other researchers lack data like these: measures of standardized learning gains do not exist in other contexts.

Riehl, Saavedra, and Urquiola demonstrate that college productivity based on learning measures produces something quite different from productivity based on earnings, especially initial earnings. Learning-based measures are more highly correlated with long-term earnings than they are with initial earnings. As in Minaya and Scott-Clayton, this suggests that learning reflects long-term value added, while initial earnings more heavily reflect students' precollege characteristics. The authors confirm this by showing that productivity measures based on initial earnings favor schools that recruit students from affluent backgrounds and whose curricula stress business and vocational fields. Productivity measures based on learning favor schools that enroll high-aptitude students (regardless of their income background) and that stress the sciences, social sciences, and arts (the traditional liberal curriculum). Riehl, Saavedra, and Urquiola's analysis thus provides a cautionary tale for performance systems based entirely on graduation rates or initial earnings—the metrics that are currently popular (see Minaya and Scott-Clayton).

So far, we have only discussed how productivity differs among institutions. However, it could potentially differ substantially by field or program within each institution. Joseph G. Altonji and Seth D. Zimmerman, in chapter 5, "The Costs of and Net Returns to College Major," analyze whether productivity within an institution differs by college major. They begin by noting that people have often thought that they have answered this question when they have simply examined the initial earnings of graduates by college major. There are a few reasons why such a simplistic exercise does not suffice. For instance, there is substantial selection into majors: students with higher college aptitude tend to major in certain fields, and their higher earnings cannot be attributed entirely to their field. Also, the relationship between initial earnings and lifetime earnings varies by major. Engineering

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

majors, for instance, have high initial earnings but subsequently experience unusually slow earnings growth.[1] However, Altonji and Zimmerman investigate a third and completely different explanation as to why initial earnings by college major are not a reliable guide to *productivity* by major: different majors cost different amounts, and costs are the denominator in any calculation of productivity.

The reason why cost differences between majors have rarely, if ever, been systematically analyzed is because even administrators often lack information on how much their school spends educating a history major, say, as opposed to a chemistry major. For want of data, administrators and researchers alike have therefore assumed costs are the same across majors. However, using uniquely detailed administrative data for all Florida public institutions, Altonji and Zimmerman show that this assumption is false: majors that are intensive in equipment, space, or faculty (especially highly paid faculty) can be dramatically more costly on a per-student basis. Examples are engineering and health sciences. The least-expensive majors require no equipment, need no dedicated space, have large classes, and have modestly paid faculty. An example is psychology. The authors show that if we consider costs, the productivity findings are very different from what we might conclude from a naive look at initial earnings. Strikingly, the ratio of initial earnings to costs is similar in majors with high earnings and high costs (such as engineering) and modest earnings and modest costs (such as public administration). The majors with the highest ratios of initial earnings to costs are ones such as business that have relatively high earnings but relatively low costs.

Few if any higher education leaders use data like that of Altonji and Zimmerman to make similarly sophisticated calculations of how productivity varies across majors. But perhaps they use less-formal, quantitative means to reach similar conclusions? If so, we might expect that they reallocate resources toward more-productive majors and away from less-productive ones. This is one of the important questions addressed Paul N. Courant and Sarah Turner in chapter 6, "Faculty Deployment in Research Universities."

In addition to the US Department of Education's administrative database that covers all research universities, Courant and Turner use internal data from two important public research universities.[2] The latter data, the likes of which are rarely available to researchers, allow them to study the productivity of individual faculty in an incredibly detailed way. For instance, they know how many students are in each class and whether it is taught by a faculty member alone or with the help of nonfaculty instructors, graduate students, and so on. Their measure of teaching productivity is novel: essentially, how

1. For evidence on these points, see Hoxby (2018).

2. The database is the Integrated Postsecondary Education Data System. It is, in fact, used to some extent by every study in this volume, a demonstration of its ubiquitous value.

You are reading copyrighted material published by University of Chicago Press. Unauthorized posting, copying, or distributing of this work except as permitted under U.S. copyright law is illegal and injures the author and publisher.

many students are produced by each dollar of faculty pay. This measure makes sense if each student generates about the same amount of revenue.[3]

Courant and Turner first demonstrate that faculty differ greatly in both their research productivity and their teaching productivity. These differences occur across universities, across fields within a university, and across faculty within a field within a university.

This is a fascinating finding, but what explains it? Here, Courant and Turner demonstrate how economic reasoning can guide hypotheses. They argue that faculty in different fields must be paid different amounts because their *outside* labor market opportunities differ. For instance, an economics or business professor's pay outside of academia would be high relative to that of a classics professor's. Moreover, the authors argue, it is very difficult for universities to reallocate either students or faculty across fields to equalize productivity. Given that universities compete for students in a market and students choose their fields based on factors that include later remuneration, universities cannot plausibly force students to major in undersubscribed fields simply to raise faculty productivity there. Universities cannot reallocate faculty easily for reasons that are both more mechanical and more economically subtle. As a mechanical matter, a professor who is expert in chemistry is not prepared to conduct research or teach courses that require expertise in history or vice versa. (Of course, a university can gradually reallocate its faculty by slowing hiring in some fields and accelerating it in others. Courant and Turner show that universities do this to some extent.) But even if a university could easily reallocate its faculty to equalize productivity fully, it has incentives not to do so but rather to protect a critical mass of expertise in all fields. A research university that failed to comprehend all fields of knowledge would have difficulty attracting philanthropic and government support.

Constrained by the labor market, the market for students, and the market for support, how can universities align faculty pay and faculty productivity? Courant and Turner show that they do this (perhaps as much as they can) by allowing larger class sizes and more nonfaculty teaching in fields where faculty are expensive. Faculty who are more productive researchers are allocated less undergraduate teaching and more time for research—compensating differentials for their not taking jobs outside of academia. On the whole, one comes away from the study with the impression that, though constrained in many and complex ways, universities maximize productivity more than one might think based on their reluctance to conduct formal analyses.

3. This measure of teaching productivity applies less well to the most selective private institutions for two reasons. First, revenue differs greatly across students because they pay more differentiated tuition, because gifts are associated much more with some students than others, and because no revenue comes from state appropriations (which tend to be made on a flat per-student basis). Second, a faculty member's influence on the world is less likely to run through mass teaching than through instructing relatively few but stellar students.

You are reading copyrighted material published by University of Chicago Press. Unauthorized posting, copying, or distributing of this work except as permitted under U.S. copyright law is illegal and injures the author and publisher.

Pieter De Vlieger, Brian Jacob, and Kevin Stange, in chapter 7, "Measuring Instructor Effectiveness in Higher Education," examine similar issues, but in a very different part of the postsecondary market: the for-profit sector where online course-taking is prevalent and institutions focus almost entirely on teaching, especially the teaching of elemental college courses. This sector also differs greatly from the research university and selective college sectors because the for-profit institutions' objectives are fairly unidimensional (profits) and because 100 percent of their revenue comes from tuition (as opposed to philanthropy or government appropriations). For-profit institutions therefore only have incentives to operate programs that attract students and that can be taught at a low-enough cost to turn a profit. Furthermore, their students tend to be intent on receiving educational credentials in order to raise their earnings as opposed to having more-complex goals, such as becoming liberally educated, earning a top professional degree (medical, for instance), or learning to conduct research. In short, we should think of the University of Phoenix, the institution the authors study, as facing very different economic incentives and constraints than research universities and selective colleges.

De Vlieger, Jacob, and Stange estimate instructor productivity in college algebra, a course required of all students in the University of Phoenix's baccalaureate programs. They observe more than 300,000 students and 2,000 instructors, a testament to the size of the institution. Using internal administrative data, the authors show that the assignment of students to teachers is virtually random once they condition on the identity of the course, section, level, and student characteristics. Thus they analyze what is essentially a randomized controlled trial and, as a result, produce highly credible results.

De Vlieger, Jacob, and Stange show that the algebra instructors vary greatly in their productivity. A 1 standard-deviation increase in their value added translates into a 0.3 standard-deviation increase in students' math skills. Variation is also great when instructors' productivity is measured by students' taking follow-up courses or earning subsequent credits. Interestingly, instructors' productivity varies more for in-person than online courses. Put another way, if students want to obtain instruction that has maximum value added, they must do it in person because the online experience suppresses variation in instructional value added.[4] This result has important implications, to which we return below.

De Vlieger, Jacob, and Stange's most striking result, from the economics perspective, is that the University of Phoenix pays these highly variant instructors exactly the same amount. Thus performance differences translate directly to large productivity differences across instructors. This is not because the institution is constrained to do so: its instructors are on short-term con-

4. Of course, if students wanted to experience unusually low instructional quality, they might also seek out in-person settings.

You are reading copyrighted material published by University of Chicago Press. Unauthorized posting, copying, or distributing of this work except as permitted under U.S. copyright law is illegal and injures the author and publisher.

tracts and are not unionized. Of course, it is possible that the institution, having learned from the study, will henceforth make pay more differentiated. Alternatively, the result suggests that the institution's profits are more purely a function of its total enrollment than they are a function of its students' success in acquiring skills or attaining credentials. If the sort of students who consider nonselective for-profit institutions do not make their enrollment choices based on the schools' record of skill production, it might make sense for the University of Phoenix to pay all instructors the same amount.

In 2006, the US Department of Education eliminated a rule that constrained postsecondary institutions to offer no more than 50 percent of their courses online. This rule had forced institutions such as the University of Phoenix to locate instructional space in metropolitan areas with sufficient population density that each space could attain minimum efficient scale. (Minimum efficient scale requires enough students taking each class and enough total classes to justify renting and managing the space.) Moreover, the rule change occurred at a time when broadband service had become available almost everywhere, even in areas of low population density.

In chapter 8, "The Competitive Effects of Online Education," David J. Deming, Michael Lovenheim, and Richard Patterson show that the 2006 rule change allowed online enrollment to expand greatly and, more specifically, into markets that had previously supported only a few postsecondary institutions. For the brick-and-mortar institutions in those markets, the change potentially constituted a major and rather sudden increase in competition. For the students in those markets, online enrollment constituted an increase in their educational options, especially with regard to price and timing flexibility. (It is not obvious that online enrollment constituted a major increase in curricular options. This is because online postsecondary programs remain, probably for technical and cost reasons, focused on fairly standard courses and credentials that were likely already available locally, even in markets with only a few brick-and-mortar institutions. The availability of online education has not yet much affected the ease with which a person can earn, say, a degree that is equivalent to one from a major research university or selective college.)

Deming, Lovenheim, and Patterson show that the increase in competition reduced enrollment at private, nonselective brick-and-mortar institutions located in areas where they had previously been one of only a few such choices. This makes sense because they are the closest substitutes for online institutions that also tend to be nonselective and that offer similar curricula.

A superficial economic analysis might then suggest that the private, nonselective brick-and-mortar schools would respond to the competition by reducing tuition. But they need not compete purely on price. Indeed, when we recall that instructors' productivity varies more for in-person classes, economics helps us anticipate what actually happened: private brick-and-mortar schools reduced class size and raised tuition. Such a response could

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

only be a market equilibrium if the brick-and-mortar and online institutions were becoming increasingly differentiated on grounds *other* than price. What seems most likely (and in accord with models from industrial organizations) is that the brick-and-mortar schools began to specialize in students who valued a fairly intimate in-person experience where, as previously shown, instructors can exercise their talent more than they can in online classes. The online institutions probably specialized in students who put greater weight on price or the flexibility of the timing of their classes.

These results remind us that as the market for (brick-and-mortar) higher education became dramatically more geographically integrated and competitive during the 20th century, institutions did not merely compete on price (tuition) but instead became differentiated on student aptitude, curriculum, and many other dimensions (Hoxby 2009). The market for postsecondary education has never been a simple market for an undifferentiated good where pure price competition prevails. It is interesting that such a statement applies to nonselective institutions focused on teaching elemental courses, not only to research universities and selective colleges.

In the most recent Beginning Postsecondary Students study, 94 percent of students who commenced their postsecondary education at a two-year public institution (community college) stated, in their first year, that their degree goal was a baccalaureate degree.[5] This would suggest that such schools' productivity ought to be evaluated, at least in part, on whether they allow students to achieve that nearly universal goal. Interestingly, almost no studies prior to Carrell and Kurlaender's, in this volume, attempt such an evaluation. This is largely because prior studies often depend on data sources that do not reliably track students as they transfer from two- to four-year colleges. Carrell and Kurlaender, instead, use remarkable administrative data that allow them to follow all California students from their high schools, to community colleges, and on to the California State University campuses (which are the destination of the vast majority of students transferring from two- to four-year colleges).

In chapter 9, "Estimating the Productivity of Community Colleges in Paving the Road to Four-Year College Success," Scott E. Carrell and Michal Kurlaender estimate each community college's productivity, where the outcomes of interest are the probability of students making a successful transfer to a four-year college (the "extensive margin") and the achievement of those students once at four-year colleges (the "intensive margin"). The data are so rich that the authors can adjust for several measures of students' incoming preparation and high school quality. They can effectively control for students' unobserved motivation and interests, normally unobservable, by controlling for the identity of each four-year college.[6] That is, they can

5. Authors' calculations based on National Center for Education Statistics (2016).
6. Of course, this strategy works only for the intensive margin estimates.

You are reading copyrighted material published by University of Chicago Press. Unauthorized posting, copying, or distributing of this work except as permitted under U.S. copyright law is illegal and injures the author and publisher.

compare two students who not only had the same high school achievement but also both transferred to Cal State-Chico having previously attended different community colleges.

Carrell and Kurlaender find that community colleges' productivity, in terms of successful transfers, differs substantially. Despite enrolling approximately the same students and relying on the same set of destination (four-year) schools, some community colleges are significantly more likely to induce a student to attain baccalaureate education. Moreover, more- and less-productive community colleges are often located fairly close together, and the less-productive ones continue to attract students.

This final result suggests that students who make up the bulk of demand for community colleges either (1) choose programs without having very much information about the program's likelihood of helping them achieve their goals or (2) (despite being informed) choose programs based on attributes that are only weakly correlated with the program's productivity. For instance, they might choose a less-productive program simply because of its proximity or the timing of its class schedule. Carrell and Kurlaender's analysis thus lends support to Hoxby's conclusion that market forces are not disciplining the productivity of nonselective institutions.

Some Immediate Takeaways

This brings us to a few takeaways for university leaders, policy makers, and researchers.

First, although our findings suggest that economics delivers powerful insights about all institutions of higher education, the market forces that drive nonselective, selective, and research institutions differ. It is not that these institutions function in disjoint markets. Rather, the market is sufficiently differentiated that, as we move around within it, the circumstances that schools face change. Students' enrollment choices appear to weigh different factors. The sources of revenue differ. The outputs valued by funders differ. There are changes in the relevant production function—whether it includes research, for example. Thus, although we are confident that economic reasoning is crucial for strong analysis of higher education, we are mindful that deep institutional knowledge is required if we are to apply economics well. It would be specious simply to transfer thinking from, say, the analysis of for-profit industries to the analysis of postsecondary education.

Accounting for selection and measuring costs in higher education are serious problems for analysis. However, there are also serious remedies available if researchers have the right data. We have been repeatedly struck by the fact that the remedies employed by the studies in this volume would have been impossible without data only recently made available. This suggests that if higher education is to learn about itself and improve, it must allow and even expand access to data for well-grounded research.

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

We recognize the problems inherent in measuring the outcomes produced by higher education, but we do not believe that they are so insurmountable that it is better to abandon the effort to measure them. Although there is a great deal to be learned by studying outcomes not measured by studies in this volume, we believe that we gained valuable knowledge from the outcomes we were able to evaluate. Moreover, we believe that many of the findings in this volume would help policy makers, postsecondary leaders, and students make more informed decisions.

Changing the Conversation

We began this introduction by describing a conversation that not only really occurred but that is fairly typical of the conversations that take place within institutions of higher education. We find that sort of conversation—in which some (but only some) costs were discussed and no benefits were discussed—frustrating. If all the participants in that conversation had read this book, would it have been more insightful?

We would argue "yes" in the following sense. Many are the proposals for improving the productivity of institutions of higher education. These range from proposals to direct most students to community colleges, to proposals to funnel students into certain majors, to proposals to move most learning online, to proposals to put all instructors on one-year contracts, and so on. Most of these proposals are based on little more than speculation. We believe that studies such as the ones in this volume show how discussions of these proposals could become reasonably grounded in evidence. In other words, the purpose of the studies in this book is *not* to prescribe productivity-enhancing policies for institutions. (Such prescribing would, in any case, violate the NBER's mission.) Rather, the studies in this volume make the case that future conversations can be informed by evidence both about benefits and about the full array of costs. In medicine, better diagnoses lead to better solutions. We believe that similar logic applies to higher education.

However, we observe that we still face formidable challenges in changing some parts—very important parts—of conversations like the one with which we began this introduction. In particular, it remains very difficult to assess the public benefits of higher education for civil society, the macro benefits for the economy, and the benefits for individuals' well-being (the "nourishment of the soul"). These benefits do not lend themselves to modern empirical research in which experimental methods feature prominently. Nevertheless, from the ancients onward, commentators have argued for the importance of such benefits of higher education. Providing empirical grounding so that these considerations can be included in conversations is a challenge for the next generation of studies.

You are reading copyrighted material published by University of Chicago Press.
Unauthorized posting, copying, or distributing of this work except as permitted under
U.S. copyright law is illegal and injures the author and publisher.

References

Altonji, J., and C. R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *Quarterly Journal of Economics* 116 (1): 315–50.

Hoxby C. 2009. "The Changing Selectivity of American Colleges." *Journal of Economic Perspectives* 23 (4): 95–118.

———. 2018. "It Is the Student Who Matters, Not the Student's Major: Towards Credibly Causal Returns to College Majors." Stanford Institute for Economic Policy Research Discussion Paper.

National Center for Education Statistics. 2016. *2012/14 Beginning Postsecondary Students Longitudinal Study*. Restricted-Use Data File, NCES 2016404. May.

You are reading copyrighted material published by University of Chicago Press. Unauthorized posting, copying, or distributing of this work except as permitted under U.S. copyright law is illegal and injures the author and publisher.