

Measuring Performance and Accountability in Higher Education: A Review Article on *Productivity in Higher Education*

Mary O'Mahony¹
King's College London

ABSTRACT

Credible and robust measurement of productivity in higher education institutions are needed to assess the accountability of the sector. This review article discusses the NBER volume *Productivity in Higher Education*. The volume contains articles estimating various facets of productivity in higher education including undergraduate outcomes from university, community college and online higher education, as well as the quality of teaching. The volume is an excellent example of combining economic reasoning with innovative data and clever analysis to yield credible conclusions. It contains important messages for the policy debate on scrutiny and accountability of Higher Education provision.

The Higher Education sector has been increasingly subject to monitoring, both across providers and for individuals within institutions. This accountability agenda is motivated by the desire to ensure improved decision making processes by the institutions themselves and by consumer and policy maker stakeholders, a laudable aim. However, such an agenda requires credible and robust measures of performance including productivity measures, but the reality has often been reliance on easy to measure indicators.

This review article discusses the volume *Productivity in Higher Education* edited by Caroline M. Hoxby from Stanford University and Kevin Stange of the University of Michigan and published in 2019 by the National Bureau of Economic Research and the University of Chicago Press. The volume employs state-of-the-art methodologies and data, and demonstrates repeatedly the dangers of using inappropriate metrics.

As a starting point for this article, it is useful to have a definition of productivity. In her chapter *The Productivity of US Post-*

¹ The author is Professor of Applied Economics at King's Business School at King's College London, and a visiting fellow at the National Institute of Economic and Social Research, London. Email: mary.omahony@kcl.ac.uk.

secondary Institutions, Caroline M. Hoxby defines in general terms the productivity of a higher education institution as the “value to society of its causal effect on outcomes (value added) divided by the cost to society of educating its students (social investment)”. All articles in the volume try to measure ‘value added’, adjusting the raw data to ensure the outcomes can be attributed to the institutions and not extraneous influences, and some also attempt to measure costs.

The volume begins with an introduction by Caroline M. Hoxby and Kevin Stange, and a first chapter that outlines parallels between measuring productivity in Health Care Services (HCS) and Higher Education (HE). Together they identify four main issues that affect attempts to measure the productivity of HE institutions: multiple outcomes, selection, the multi-product nature of the institutions, and attribution. The introduction identifies a fifth challenge, the public nature of benefits, which I discuss further in my concluding remarks. This review starts with outlining the main measurement challenges and how the chapters contribute to dealing with them. I then present a summary and some remarks on the main results by chapter, followed by a brief discussion on data requirements. The review ends with some concluding remarks.

Issues in Measuring Productivity in Higher Education Institutions

Multiple Outcomes

Douglas Staiger in his Chapter *What Health Care Teaches us about Measuring*

Productivity in Higher Education, argues that multiple outcomes plague measurement in HCS and that there is no easy fix for how multiple outcomes should be weighted. He warns against targeting specific outcomes as this can lead to unintended consequences for other outcomes not subject to the same level of public scrutiny. Similarly, there are many outcomes from HE, including learning outcomes, post-graduation earnings, employment, innovations and public service.

All chapters in the volume deal with multiple outcomes to a certain extent. This issue is addressed in some detail in the second chapter by Caroline M. Hoxby, where she produces productivity estimates for a number of outcomes. These include private earnings, public service measured by the difference between salaries in the private and public sectors for the same occupation, and innovation measured by R&D spend in employing firms. She shows that relative productivity varies significantly across these three outcome measures (see further discussion below). This highlights the dangers of examining one particular outcome, consistent with the warning by Staiger in his chapter.

Veronica Minaya and Judith Scott-Clayton, in their chapter *Labor Market Outcomes and Postsecondary Accountability*, investigate in some detail different outcomes, including earnings, full-time, full-year employment as a measure of employment stability, employment in social service sectors, per cent ever claiming unemployment since graduation and degree completion measures.

In Chapter 4, *An Approximation to College value Added in Two Dimensions*,

Evan Riehl, Juan E. Saavedra and Miguel Urquiola attempt to measure learning as well as financial outcomes, using results from a common exit exam in Colombia. The articles in this volume show that comparisons across institutions are very sensitive to which outcomes are chosen and how they are measured. This is an important finding in a world where education, health and public service providers are increasingly being monitored by policy makers. Choosing easy to measure outcomes, such as earnings immediately after graduation, as monitoring tools may poorly reflect true differences in productivity across institutions and may encourage gaming with longer term adverse unintended consequences. Of course, researchers have been warning about this aspect of accountability exercises for decades, but this volume presents very clear evidence that this warning should be taken seriously.

Selection

Selection is the issue that has most perplexed researchers in evaluating the performance of education institutions, both schools and HE. Put simply, the post-graduation earnings of a student from Harvard University is likely to be of an order of magnitude higher than a student from a non-selective school, but much of these additional earnings are likely to reflect the innate ability of the student as well as their family and social backgrounds. Indeed, selection is also an issue in measuring HCS, as patients do not sort randomly to hospitals. In the health case, much research has been devoted to applying risk adjustment factors to raw data and Douglas Staiger

suggests that adjusting for selection should be no more difficult for HE. In fact, he argues that it might be easier, as many universities have explicit admissions criteria that should permit identification of student characteristics, such as prior achievements or family background. However, there are many unobservables, such as ability, which might make the selection issue more difficult to deal with for HE. Selection issues feature in all the articles in the volume.

In Chapter 2, Hoxby uses an approach that attempts to mimic randomization while dealing with lack of overlap or common support. The latter is the problem that the ‘ability’ distance between highly selective and non-selective schools means there are no usable overlaps of students applying to both that could be compared. Instead she compares groups of schools adjacent to each other, using statistical paired comparison methods (PCM). Quasi-random experiments have both horizontal and vertical components. Horizontal selection, for example, selection due to the geographic location of schools, is easy to deal with by comparing students with the same application credentials choosing between two equally selective schools.

To take account of vertical selection, Hoxby considers pairs of students with the same application credentials applying to schools that are not equally selective. She uses the fact that in all admissions processes there are students who are well above the admission thresholds who are immediately accepted, students who are rejected outright and a group in the middle who can be given an offer depending on how many residual places are available. The offers in this ‘on-the-bubble’ range can be

thought of as admissions officers flipping a coin, since there are no obvious credential differences between students in this range. Hoxby uses this middle range in adjusting for selection, and shows that this has a large impact on relative, across institution, productivity measures.

Chapter 7 by Pieter De Vlieger, Brian Jacob and Kevin Stange, *Measuring Instructor Effectiveness in Higher Education*, investigates a setting that effectively amounts to a randomised control trial where the assignment by students to instructors is random after conditioning on a number of course and student characteristics. Many chapters take a more conventional approach of using control variables to adjust for selection, as often data are not in a form conducive to mirroring random experiments. For example, in Chapter 3, Minaya and Scott-Clayton rely on regression based adjustment factors to address selection issues. Nevertheless, dealing with selection issues is a central feature of all chapters and the volume is an exemplar of how to address these issues to produce credible results

Attribution and Multi-Products

The HE sector produces multiple products across teaching programmes (undergraduate, post graduate taught, professional and doctoral programmes) and institutions vary enormously in their teaching focus. In addition HE institutions produce research and hospitality services, although these are not the focus of the volume. However chapter 6 by Paul N. Courant and Sarah Turner, *Faculty Deployment in Research Universities* examines linkages be-

tween research and teaching, both volumes and types of teaching – undergraduate versus postgraduate.

The attribution issue has always been an issue in HCS, as patients receive treatments by multiple hospitals and providers. The same is true for HE, as some students progress from two year college to four year undergraduate to postgraduate degrees, and especially in the US case, students frequently change institutions even within their primary degree. In HCS this is often dealt with by identifying health episodes that span multiple providers. In theory this could also be done for HE by tracking students across education providers. Similarly, taking from the health example, it may be beneficial to attribute outcomes to the first institution a student attends, as all other choices depend on this.

The attribution issue is dealt with in Hoxby's chapter by using lifetime earnings and all social investments. Many chapters opt to deal with these issues only implicitly, by focusing on undergraduate education and often restricting the sample, for example, to those who enter the labour market immediately after graduation. The exception is the final chapter by Scott E. Carrell and Michal Kurlaender, *Estimating the Productivity of Community Colleges in Paving the Road to Four-Year College Success*, which examines transfers from two year to four year Colleges.

Summary of Main Findings

Outcomes and Productivity in HE Institutions

In Chapter 2 Hoxby finds that, among

selective schools, value added lifetime earnings rise with the degree of selectivity of schools, but so does social investment in terms of educational resources and students' capacity to use those resources based on their ability, family backgrounds, etc. Interestingly, total social investment rises more steeply than tuition fees. This is partly due to much higher income from donations from alumni and other philanthropists as well as income from endowments, but also reflects that students in the most selective schools enrol in more curricular units and are less likely to drop out. The net impact is that productivity is relatively flat for this group of schools. She interprets this as saying that market forces operate – this part of the HE market competes both for the best students and for faculty and other resources. The implication is that taking one dollar from the most selective schools and giving it to one whose entry requirements are somewhat lower would not raise productivity of the HE sector overall. The allocation of resources to students roughly correspond to their ability to benefit from them.

This is, to my mind, a surprising result, given that, as acknowledged by Hoxby, there are many facets of HE, such as funding by taxpayers and information asymmetries, which imply market forces do not function perfectly. Partly this result is driven by the value added approach employed in the article, and highlights the need to take account of selection issues. However, the Figures in the article suggest this is not the major contributing factor. It appears that the US HE sector for selective schools is mirroring a competitive system.

When non-selective schools are com-

pared to selective ones, however, the former have much lower productivity, and more within group variability. Here horizontal differentiation is probably more important, with students deciding according to geographic location or other amenities. Here Hoxby argues that market forces are less likely to operate as students have less information on which to base their decisions and fees are frequently paid by third parties.

When examining the other outcomes, Hoxby shows that for education contributions to public service, productivity rises with selectivity but the dispersion is much greater than that based on private financial returns. Some very selective schools have much greater productivity in contributions to public service, and this may reflect the ethos of the institutions. Market forces are unlikely to play a major role here. Finally, the innovation productivity measures show a very steep upward slope for the most selective schools. Since innovation often spills over to other workers there is no reason to expect market forces to operate. In addition the most selective schools are also the most research intensive, and so faculty are more likely to engage in research led teaching, and undergraduate programmes might be more research oriented.

Overall, Hoxby's chapter paints a picture of undergraduate education, that generates considerable benefits relative to social investments in selective institutions. For non-selective institutions, this conclusion is not so clear cut but the comparison here should arguably be relative to alternatives such as on the job training. This chapter provides a comprehensive overview of performance in HE institutions, using

administrative data from tax returns linked to official reports of these institutions from the US Department of Education. It illustrates what can be achieved by combining high quality and comprehensive data sources with extensive knowledge of education systems and is a must read for anyone interested in evaluating performance in the Higher Education sector. This chapter cannot, of course, cover all aspects of HE provision, and many of the details are covered in the remaining chapters.

The context for Chapter 3 by Minaya and Scott-Clayton is the drive towards performance funding in many states. Often these rely on crude measures such as completion rates or earnings very soon after graduation. Using data for the state of Ohio, and administrative unemployment insurance records, the authors measure outcomes four years after graduation, although they include some sensitivity analysis to using shorter or longer periods, and undertake separate analyses for 4-year and 2-year degree awarding institutions. They show that there is a high degree of variability of the relative performance across institutions depending on the outcome measures – earnings, employment based measure or completion rates. Most outcome measures are highly positively correlated with the exception of degree completions, especially for the two year degree sector. Statistical compositional adjustments are more important for early outcomes, in particular for the four year programmes, but become less so as time goes on.

The general conclusion is that completion rates and early earnings outcomes, preferred by policy makers, are not good indicators of longer-term outcomes such

as earnings or employability later in life. The choice of metric and the length of the follow-up matter greatly, and the longer the latter the less selection based on family background or wealth matter and the more outcomes are based on skills acquired through HE.

Riehl, Saavedra and Urquiloa in Chapter 4 use unique data for Colombia that allows the estimation of direct learning outcomes. All graduating students sit a common exit exam, which can be divided into field specific tests and reading and English that are common components across all exit exams and which partly match to entrance exams. They show that the correlation between earnings and learning outcomes are not strong, especially after adjusting for individual characteristics. Selective public institutions appear more favourable when productivity is measured in terms of learning while the best private colleges perform relatively better in terms of earnings. Consistent with the previous chapter, the authors show that results can be very different depending on how soon after graduation earnings are measured. Learning outcomes are more highly correlated with later earnings, reflecting longer term value added, whereas short term earnings are more influenced by student characteristics.

The authors also show that there are variations across fields, with engineering and business degrees showing a more favourable productivity performance in terms of earnings, and subjects such as arts and humanities performing better in terms of learning outcomes. As with the previous chapter, the authors conclude with a discussion of accountability issues, and the need for these to be designed to incorporate

the findings that the choice of outcomes and the time periods over which they are measured are likely to lead to very different rankings of HE institutions.

The chapter by Altonji and Zimmermann delves further into productivity when dividing by College Major, examining both earnings and costs at a level of detail not available to date in the literature. Using data for Florida universities they show that there is significant variation in earnings across majors, which is well known, but also large variation across costs, both per graduate and per credit. Some subject areas such as engineering, with high earnings and high costs, have similar productivity to majors with low earnings and low costs such as social sciences. The results suggest that variations in costs are economically significant. This is important information for policy since it is clear that a marginal dollar spent does not have equal value in all fields. Differences can be justified if some fields such as high cost but low earnings physical sciences generate externalities, but the authors are sceptical that spillovers could be sufficiently large to account for the differences they observe. The authors also highlight changes over time, and show a reduction in costs from 1999 to 2013 in aggregate and for most majors. An interesting observation is that this is partly driven by changes in the composition of instructional resources from full-time permanent faculty to more temporary adjunct faculty. Although beyond the scope of their study, the authors point to some prior research that this may be at the cost of lower learning outcomes.

Faculty and Instructional Resources

The last point leads neatly into the analysis in chapter 6 by Courant and Turner, *Faculty Deployment in Research Universities*. Here the focus is on research intensive universities whose faculty both teach and research. The analysis is based on two public universities where the authors can access very detailed information on teaching loads and salaries. They show that there is enormous variation in salaries across fields - salary differentials largely reflect outside opportunities. However, 'costs per student' are negatively correlated with salaries. Therefore, universities vary the organization of teaching across departments to reflect input costs, leading to greater productivity. When examining interdepartmental allocations of teaching the authors show that salaries are negatively correlated with teaching load and positively with quality teaching (postgraduates versus undergraduates). The allocation of resources ensures those who are good at research make best use of their talents. Over time the authors show that salaries at research intensive universities have increased significantly, reflecting an increase in the price of research relative to teaching. The authors' results suggest that universities have responded by allocating teaching more efficiently, through larger class sizes in fields where faculty are relatively more expensive.

Instructor productivity is also the subject of chapter 7 by De Vlieger, Jacob and Stange, but in a very different setting, the for-profit HE sector. Here teaching is the only activity, and takes place both face to face and online. Based on exam results

and progression across courses, the authors show that there is huge variation in instructor effectiveness. To put this in context, the variations across instructors is greater than that for outcomes for students aged 35 relative to aged 25 or those whose GPA is 2.00 relative to those whose GPA is 3.00. This is a large difference and one that the authors find is neither correlated with student's direct evaluations of their instructors or with pay, which if it varies at all, does so on the basis of seniority. This large difference in productivity of instructors suggests scope to improve student performance through personnel policies that cover how instructors are hired, retained, motivated and developed.

Online Education

Chapter 8 by David Deming, Michael Lovenheim and Richard Patterson, *The Competitive Effects of Online Education*, attempts to answer the question of how online education affects the market for HE. Online education increases choices for students, especially for those who had few prior options due to local monopolies in the non-selective or for-profit sectors. The authors exploit a change in the law in 2006 that eliminated a requirement that no more than 50 per cent of courses could be distance learning for schools that received federal aid. In a generalised difference in difference approach, the authors find that, after the expansion of online degree programs, less competitive markets experience declines in enrollments and that the impacts are concentrated in private institutions. They find little evidence of a negative impact on tuition fees in the private

sector – public institutions receive heavy subsidies so are not considered – and in fact a significant positive impact for the private four year institutions. Therefore they provide evidence that these institutions do not compete in price. Instead they find that online competition increases instructional expenditures per student, and more so for public institutions who probably try to compete on the quality of the education offering, reducing class sizes and offering a better experience to students who prefer an in-person experience.

The results are robust to using internet penetration rather than degree of market concentration as their difference measure. Online degrees are often seen as poor quality in terms of learning with lower completion rates and worse labour market outcomes than traditional face to face teaching. The authors highlight an important positive from these courses, arguing that they can raise productivity and innovation in competing institutions. It will be interesting to see in the future if the move to more online delivery in all HE sectors due to the COVID-19 crisis will crowd out specialist online programs or lead to more students enrolling in them. The former could occur if bricks and mortar HE institutions are sufficiently innovative in how they blend learning to incorporate some online material but the latter may be the outcome if students do not perceive a sizeable difference in their learning experiences.

Transfer between HE Institutions

The final chapter by Carrell and Kurlaender investigates variations in the degree to which two year community col-

leges facilitate transfer to four year institutions (the extensive margin) and the performance of students given that they transfer (the intensive margin). The article uses rich data for California that allows controlling for student and institution characteristics. The results show very wide variation in transfer rates and in post transfer performance. The latter is measured by GPA in the first term at the four year college, the probability of persisting to year two and of graduating and length of time to graduation. The authors also present some evidence of a positive correlation between community colleges that are successful in transferring students and those where student subsequently perform well. Observable characteristics of the more successful colleges include their size, their geographical proximity to four year College campuses and a higher ratio of female faculty. The variation in productivity might reflect students preferences over amenities or location, or lack of information by students. The results are consistent with those found by Hoxby that market forces do not appear to equalise productivity for non-selective school.

Measurement and Data

The volume produced many interesting findings as summarised above. It also illustrates the powerful and robust conclusions that can be drawn when researchers have access to very good quality data. In all chapters, the analysis is based on administrative data sources, that allow tracking and matching individuals or institutions across many dimensions. This permits much clearer identi-

cation of the research question of interest, while controlling for important extraneous influences, than more traditional aggregate data sources. From across the pond, researchers can only envy the richness of the data available. That is not to say that such data are not collected in the UK or other European countries – universities collect enormous amounts of information on their students and faculty and the tax authorities collect information on graduate salaries and benefits. The main issue is difficulty of access for researchers. In Europe there are drives towards allowing greater access, but at a slow pace. However, much more work is needed in this respect, even in the United States, a point made by Hoxby and Stange in their introductory chapter.

Conclusion

This book has appeared at an opportune time, although the authors when writing were unaware of the momentous changes that inevitably will take place in Higher Education around the World, as a fallout from the coronavirus crisis. HE institutions face enormous challenges in the near future, in delivering education in safe environments and moving quickly to supplement their traditional methods using on-line resources, as well as likely declines in the numbers of international students. It becomes more imperative than ever to consider the productivity of these institutions.

In the opening paragraph of the introduction, the editors motivate the research by recanting a tale of HE administrators focusing solely on cost rather than benefits, which they often see as unmeasurable. This resonates with any researcher who has

attempted to disseminate to policy makers/administrators the results of measuring productivity in public services or hard to measure sectors such as HE. It is very difficult to convince. Indeed, in his chapter Staiger emphasises that stakeholder buy-in is important for policy-makers, consumers and tax-payers to take seriously the productivity measures. This volume does an excellent job of dealing with the measurement issues and shows how much can be achieved by using administrative data sources coupled with knowledge of the pitfalls of using raw data and combined with sound economic reasoning.

Where the volume falls short is in mea-

suring the wider public benefits from HE. There is a vast literature on correlations linking university education to individual outcomes such as better health and more stable family structures. Similarly, there are correlations with societal outcomes such as lower crime rates and greater civic engagement. It is much more difficult to demonstrate a causal relationship, that separates the impacts on individuals and society of higher education from other factors. *Productivity in Higher Education* is, therefore, just a starting point for a wider research agenda on the value of education. But it is a very good starting point and well worth a read.