FISCAL STUDIES, vol. 39, no. 2, pp. 241–264 (2018) 0143-5671

Class Size at University*

GERVAS HUXLEY,† JENNIFER MAYO,‡ MIKE W. PEACEY§
and MADDY RICHARDSON◊

†*University of Bristol*
(gervas.huxley@bristol.ac.uk)
‡*University of Michigan*
(jenmayo@umich.edu)
§*New College of the Humanities*
(mike.peacey@nchlondon.ac.uk)
◊*University of Cambridge*
(mr701@cam.ac.uk)

Abstract

An effective higher education market should increase educational standards. For universities to fulfil this role, students need reliable information about the teaching on offer at different universities, but no such data are currently available. We define a measure of teaching that weights contact hours by their intensity and collect a new data set that allows comparison of teaching across universities and across three departments.

No two universities offer identical teaching. There is large variation in contact hours and even larger variation in teaching intensity, across both universities and departments. We combine our data with existing data to investigate the relationship that teaching has with university and student

* Submitted October 2015.

This project benefited from funding from the Centre for Markets and Public Organisation (CMPO) through ESRC grant number ES/H005331/1. The authors are grateful for comments from seminar participants at New College of the Humanities, ISER (Essex), LEER (Leuven) and AEDE (Madrid). They are also grateful to Gill Clarke (for bringing to their attention the appendices to the Robbins Report) and for comments from Jack Britton, Olga Gdula, Ellen Greaves, Nick Hillman, Maria Mullens-Burgess, Richard Murphy, Helen Simpson, Sarah Smith, David Willetts, David Winter and Gill Wyness.

Keywords: class size, student choice, Teaching Excellence Framework (TEF), teaching intensity.

JEL classification numbers: I23, I28.

characteristics. We find that how much teaching students receive is uncorrelated with tuition fee; that teaching has little predictive power in explaining student satisfaction; and that physics students consistently receive more teaching than either economics or history students.

Policy points

- In contrast to many dimensions of quality (for example, research reputation), it is difficult for prospective students to benchmark universities in terms of the teaching provided. This results in an important informational market failure.
- We propose an input-based metric that weights contact hours by teaching intensity and makes it possible to compare teaching delivered in different ways at different universities. Many universities already hold the administrative data required to construct this metric at the subject level.
- There is large variation in the teaching received both between and within subjects. Students receive 2.3 times more teaching in physics than in history and 2.9 times more than in economics. The ratios of maximum to minimum teaching provided across universities are 21.6, 6.4 and 25.8 for economics, history and physics respectively.
- We argue that our proposed measure can complement existing metrics to increase transparency and improve student choice.

> A competitive and dynamic higher education sector needs students who actively and regularly challenge universities to provide teaching excellence and value for money.
>
> Department for Business, Innovation and Skills, 2016

I. Introduction

In this paper, we examine the teaching arrangements at UK universities in more detail than any study since the 1963 Robbins Report. We find that whilst contact hours have changed very little, class size has increased very substantially in the last 50 years. We also find large variation in how much teaching is received by students both between and within subjects.

One of the objectives of a university is to attract and retain high-quality students.[1] To achieve this, universities must compete in terms of characteristics that students care about, and this requires students to have readily comparable information that helps them decide where and what to study. There are many characteristics that students might base their choice on: reputation (especially of research); infrastructure (sports facilities, accommodation, etc.); employment prospects; student satisfaction survey scores; and high-quality

[1] De Fraja and Valbonesi, 2012.

teaching. If teaching is unobservable, students can do no better than to make decisions based on characteristics that they can observe. Providing meaningful information on teaching is not straightforward. First, there are no agreed definitions of a lecture, tutorial, seminar or laboratory. Second, even if agreement can be reached on what is meant by these terms, universities offer very different bundles, which adds to the problem of inter-university comparison.

In the UK, the Higher Education Statistics Agency (HESA) collects administrative data from higher education institutions across a number of areas. For example, the HESA student record comprises student-level data on entry qualifications and personal characteristics, course-level data, funding, and qualifications awarded. However, there is a paucity of data on how teaching is delivered, which prevents any comparison of class size or contact hours across universities. Starting in 2006, the annual Student Academic Experience Survey undertaken by the Higher Education Policy Institute (HEPI) is completed by approximately 15,000 full-time undergraduates.[2] This survey includes a question about the contact hours students experience in different-sized classes. However, due to the way in which class sizes are grouped, most variation in class size is unobservable. This means the survey is of only modest help to students making choices about teaching. The situation in other countries is similar: as far as we know, no country uses administrative data to provide students with reliable information on how teaching is delivered at different universities. We therefore define a measure of 'teaching intensity', which incorporates information on numbers of students per lecture, class and practical across all modules within an undergraduate degree programme, thus providing a single summary measure to capture teaching inputs. Using the rights contained in the Freedom of Information Act (2000) (FoI Act) in the UK, we sent an identical questionnaire to 99 universities asking for detailed information about the teaching delivered on three courses: Economics BSc (L100), History BA (V100) and Physics BSc (F300).

Concerns about performance of higher education institutions are increasingly widespread: for example, 'No actors in the system are primarily interested in undergraduate student academic growth'.[3] In the UK, this has led to legislation that will result in the most fundamental shake-up in the architecture and governance of higher education in a generation.[4] The main policy implication will be the introduction of a Teaching Excellence Framework (TEF) to complement the existing Research Excellence Framework (REF). In our timely research, we demonstrate the feasibility of collecting data on contact hours and teaching intensity and show how this information can

[2]Higher Education Policy Institute, 2016.
[3]Arum and Roksa, 2011.
[4]Department for Business, Innovation and Skills, 2016.

be used to increase transparency and improve the functioning of the higher education market.

In Section II, we present a brief literature review of the determinants of the quality of teaching in higher education. In Section III, we describe the existing data. In Section IV, we explain the process by which we collected our data. In Section V, we present summary statistics of our data to describe the current state of teaching in higher education. In Section VI, we explain the problems that arise in interpreting these data and we introduce our teaching intensity metric. In Section VII, we combine our data with existing publicly available data to analyse the determinants of both teaching and student satisfaction scores. In Section VIII, we argue that one explanation of our findings could be the existence of informational market failures in the UK higher education market, which specifically relate to teaching.

II. Literature

The influential *Dimensions of Quality*[5] lists the characteristics that determine students' performance and learning gains: class size, cohort size, extent of close contact with teachers, teacher quality, the extent and timing of feedback on assignments and the extent of collaborative learning.[6]

There is a small literature on class size in higher education.[7] The results of literature on class size are, like the related literature on class size in schools, ambiguous. Both Krueger (2003) and Schanzenbach (2010) have argued, in the context of schools, that the results of research on class size are largely determined by the quality of the evidence. Randomised trials (which provide a valid counterfactual) invariably find that class size is important. However, the relevant policy question concerning class size is complex because smaller classes frequently involve offsetting changes.

Bandiera, Larcinese and Rasul (2010) use administrative data on individual students from a leading UK university who are enrolled in one-year MSc programmes for the academic years 1999–2000 to 2003–04. They find that the effect of class size on student grades is both significant and non-linear. The effect on performance of moving from a class with between 1 and 19 students to a class with 20–33 students is large and negative. Increases in class size above 33 have no significant effect on performance. The authors also find that class size and ability are complementary: the highest-ability students benefit most from a reduction in class size.

[5] Gibbs, 2010.

[6] David Willetts, when he was Minister of State for Universities and Science, wrote the foreword to Gibbs (2012) and cited his work in several speeches.

[7] For example, Martins and Walker (2006), Bandiera, Larcinese and Rasul (2010), Monks and Schmidt (2011) and Sapelli and Illanes (2016).

In the context of teaching undergraduate economics, Becker and Powers (2001), Arias and Walker (2004) and Kokkelenberg, Dillon and Christy (2008) all find that large classes are associated with negative outcomes, such as higher dropout rates for weak students, lower exam performance, less progression to advanced-level courses and less favourable student evaluations.[8] Class size is increasingly seen as a critical pedagogical variable that can be at least as important as different teaching methods, in part because it has implications for the choice of teaching method. Large classes are highly correlated with exclusive reliance on lectures and multiple-choice exams. Moreover, Allgood, Walstad and Siegfried (2015) observe that 'most studies do not account for these differences when evaluating the role of class size on student outcomes'.

Surveys of student satisfaction are an important tool for measuring outcomes in higher education.[9] In the UK, it has been proposed that a university's benchmarked score in the National Student Survey (NSS) will partly determine whether it is permitted to raise its undergraduate fee.[10] The motivation for introducing the NSS was to inform and influence student choice.[11] However, Gibbons, Neumayer and Perkins (2015) find that the additional information provided by the NSS 'has only a small impact on the choices of students'. They find that improvements that move the NSS score from the median to the 95th percentile across courses would generate a 2.3 per cent increase in applicant numbers. Cheng and Marsh (2010) conclude that use of the NSS to make comparisons across the same subject at different universities, or different subjects at the same university, must be very carefully interpreted: 'the onus is on NSS advocates to demonstrate their construct validity in relation to ways in which they are actually used as well as ways they are intended to be used'.

Since the majority of students only experience a single university, evaluations are absolute rather than relative. Moreover, it has been shown that differences in student satisfaction scores are correlated with student characteristics (for example, age, gender and previous schooling).[12] Similar arguments are made by Brown et al. (2015): 'If students are evaluating their own universities in the context of possibly incorrect beliefs about what happens at other universities, satisfaction with provision may be influenced by factors other than objective quality of educational experience'.

[8] For a survey on teaching undergraduate economics, see Allgood, Walstad and Siegfried (2015).
[9] For an econometric analysis of the UK National Student Survey, see Lenton (2015).
[10] Department for Education, 2016.
[11] Richardson, Slater and Wilson, 2007.
[12] See Badri et al. (2006) and Huybers, Louviere and Islam (2015).

III. Existing data

1. The Robbins Report (1963)

The 1963 report of the Committee on Higher Education ('the Robbins Report') was commissioned by the government to make recommendations about the future of higher education. To assist in its deliberations, the committee commissioned a series of surveys, including studies of the student experience. Detailed time sheets were completed by 4,224 undergraduate students (1 in 22 of the student population at the time), setting out how they used their time in February 1962.[13] The results of these surveys were published in a number of appendices to the report. These appendices provide descriptive statistics on almost every aspect of undergraduate teaching and learning in 1963, including contact hours, tutorials, lectures, feedback and the extent to which PhD students were used in teaching.

Figures for class size in the Robbins Report are broken down into lectures, seminars, discussion periods and tutorials.[14] Considering both tutorials and seminars, the Robbins Report estimates an average class size of 4.2. Approximately 50 per cent of lectures were attended by fewer than 20 students.[15] In 1963, students complained about the quality of their education, but the majority of such complaints were focused on demands for 'Oxbridge-style' tutorials. In Appendix A online, we investigate how contact hours and class size have changed since 1963; this is the first time such a comparison has been made.

One of the recommendations in the Robbins Report concerned written work: '121: Every student should be regularly set written work, which should be returned and discussed with [the student]'. At the time, this condition was largely met: 70 per cent of students in the humanities and 65 per cent in social science received written comments and discussed their work with tutors. The overall impression is that students were satisfied with both the quality and quantity of feedback.

The report finds that PhD students did some of the teaching[16] – but far less than today. PhD students were used in science subjects to take practicals (but not, by and large, to teach), and they were used hardly at all in arts, humanities and social science.

[13]These data are presented in the Hale Report (Hale and Tattersall, 1964) and in appendices II and III of the Robbins Report (Robbins, 1963).

[14]These were defined by the group size: tutorials (1–4 students), small seminars (5–9 students), large seminars (at least 10 students) and lectures.

[15]Robbins, 1963, appendix III, page 73.

[16]Robbins, 1963, appendix II B.

2. HEPI data (2013)

The closest parallel with the Robbins Report data is the student survey conducted by HEPI and the Higher Education Academy. The annual survey of 15,000 UK undergraduates, which has been conducted most years since 2006, was described by Hillman (2015), Director of HEPI, as ‘a helicopter with a telephoto lens hovering over institutions to find out what students are really up to’. Amongst other things, the survey includes questions on class size and contact hours.

According to the survey’s data, average weekly scheduled contact time across all institutions and subjects is 13 hours and 12 minutes. There were considerable differences in the amount of contact time between subjects: for example, just over 17 hours for physical sciences versus less than 9 for historical and philosophical studies. However, the group sizes used in the questionnaire make it very difficult to observe differences in class size. In Appendix A online, we compare these results with the data we have collected.

IV. Administrative data collection

The increase in 2012–13 of tuition fees to £9,000 per year, combined with an absence of existing data, provided the motivation for this paper. We collected administrative data for this academic year directly from universities using the rights contained in the Freedom of Information Act. The FoI Act created a right to request certain information held by public authorities, subject to certain exemptions.[17] For the purpose of the FoI Act, universities are generally regarded as public authorities.[18]

Our FoI request (reproduced in Appendix B online) was sent to every university that offered at least one of three undergraduate degree programmes: L100 (economics), V100 (history) and F300 (physics). In total, we contacted 99 universities and asked for data on those of the three they offered. This ensured that we had data from a subject in each of social science, the humanities and STEM,[19] whilst keeping the request to a size that could be met by the universities under the FoI Act.

We requested module-level data on the quantity and nature of contact hours, as well as information about the employment contracts of the teachers. This

[17]In those cases where universities were unable to comply with our request, they usually cited one or more of three exceptions: Section 12, which states that requests must take less than 18 hours to complete; Section 44, that requests must not reveal information preserved under the Data Protection Act (1998); and Section 43, that requests must not prejudice a party’s commercial interests (subject to a public-interest test).

[18]This is because the majority of universities in the UK are government financed. Private universities do not have to comply with the FoI Act.

[19]STEM stands for science, technology, engineering and mathematics.

provided us with contact hours, broken down into lectures and classes.[20] By combining this information with module enrolment, we were able to calculate (amongst other things) mean class size.

Of the 99 universities we contacted, four either were not running the degree programme in 2012–13, were exempt from the FoI Act or did not reply to our initial request. A further four universities refused to provide data and 24 provided us with data we were unable to use.[21] Our final data set contains 67 universities:[22] 59 of these provided us with the data following the initial request and eight agreed to provide the data requested following our modified FoI request.[23]

In Appendix C online, we compare the characteristics of the universities that did and did not provide the data we asked for.[24] For each of the three subjects, our sample appears to be representative. Although we do not suppose that history, economics and physics are fully representative of all subjects taught in the UK, the variation in teaching found in all three almost certainly arises in other subjects.

V. Summary statistics

In this section, we present descriptive statistics of the data we collected and show some correlations between variables. These summary statistics show the current variability in teaching arrangements across universities, which is at present unobservable to students. Table 1 shows the mean and standard deviation (SD) for the main variables in our data set.[25]

Cohort size is calculated for each year group and is assumed to be the number of students who attend the largest module delivered within the degree programme.[26] There is large variation: a lecture given to the smallest of these cohorts will feel more like a class.

[20]We defined a 'class' as any teaching where students are divided into groups smaller than the number enrolled in lectures, but which is not classified as a 'practical' (see Appendix B). In practice, universities usually call the groups either tutorials or seminars.

[21]Generally, this took the form of universities not providing us with enough information to calculate class size.

[22]The universities are listed in Appendix F online.

[23]These universities argued that collecting the information on staff contract types would exceed the 18-hour limit specified in the FoI Act. The modified request did not ask for this breakdown (see Appendix G online).

[24]We omitted Oxford and Cambridge from this study because their teaching arrangements are so different. We are currently collecting data from each of the colleges at each of these universities in a subsequent work.

[25]The average proportion of classes taught by hourly-paid staff could not be calculated for the universities that only responded to the modified FoI request.

[26]This definition does not correspond to the number of students enrolled on each of the three programmes. For example, a first-year microeconomics course might also be taken by students who are enrolled on L103 (economics with management). We include such students because we believe what matters is the size of teaching groups.

TABLE 1
Summary statistics

	Economics, mean (SD)	*History, mean (SD)*	*Physics, mean (SD)*
Cohort size (per year)	215.83	104.11	110.46
	(170.34)	(67.49)	(79.06)
Proportion of modules that offer classes	0.84	0.53	0.37
	(0.24)	(0.39)	(0.37)
Average lecture size	111.93	41.19	76.52
	(115.19)	(41.66)	(73.88)
Average class size	19.89	15.92	21.50
	(12.04)	(8.38)	(12.85)
Average practical size	66.48	36.24	25.09
	(103.02)	(42.73)	(21.76)
Total hours of lectures (per year)	169.77	148.07	201.28
	(104.61)	(73.61)	(119.76)
Total hours of classes (per year)	68.91	61.39	46.17
	(45.59)	(68.34)	(76.47)
Total hours of practicals (per year)	6.38	2.15	68.38
	(12.52)	(8.59)	(70.67)
Proportion of classes that are 'small' (9 or fewer)	0.07	0.11	0.05
	(0.15)	(0.15)	(0.06)
Proportion of classes taught by hourly-paid staff	0.16	0.10	0.15
	(0.20)	(0.12)	(0.29)
No. of observations	95	133	79

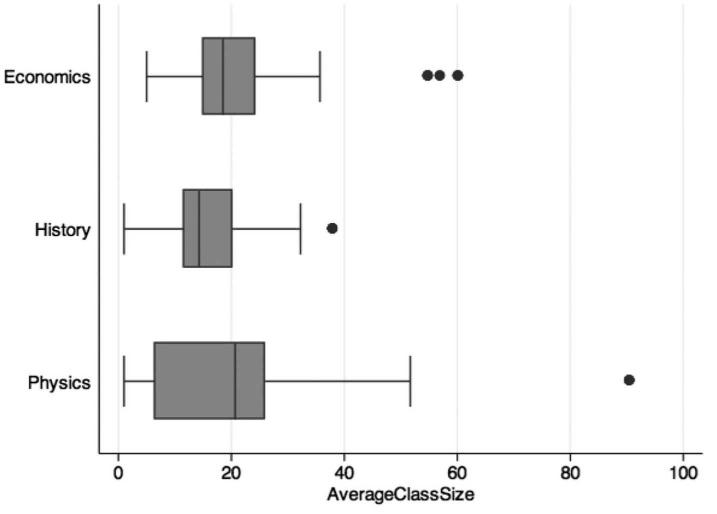
Some modules deliver all their teaching via lectures and practicals. Thirty-seven per cent of physics modules deliver their teaching via classes compared with 84 per cent of economics modules. Modules that offer no classes are excluded from our calculation of average class size.[27] The mean average class size is 15.9 for history, 19.9 for economics and 21.5 for physics. Once again, the variation in class size is large; this can be seen in Figure 1. The variable 'Proportion of classes that are small' is the proportion of teaching provided in classes of at most nine students and is calculated as a proportion of classroom teaching. History students are taught in classes of at most nine the most often.

The variable 'Proportion of classes taught by hourly-paid staff' is an estimate of the amount of teaching undertaken by PhD students. Physics and economics departments make more use of staff on these contracts (around 16 per cent of classes) than history departments (10 per cent of classes).

[27]Lectures (and practicals) are also excluded from class size calculations.

FIGURE 1

Average class size by subject



Note: In these box and whisker plots, the ends of the box show the first and third quartiles, the line inside the box shows the median and the ends of the whiskers indicate the variability outside the upper and lower quartiles. Outliers are shown by the dots.

Table 2 presents the data on contact hours, split by group size for (a) lectures, (b) classes and (c) practicals. These group sizes are inevitably arbitrary and we have chosen ones that may help the reader gain a picture of the distribution. The data in this table illustrate how misleading it is to give students information on contact hours while providing no, or uninformative, information on group sizes.

Figure 2 presents the relationship between average class size and total hours of classroom teaching, where each data point is at the year–degree level. There is a small positive correlation for history: students taught in large classes are partially compensated with additional contact hours. However, there is no such compensation in economics, and in physics the relationship is actually negative (although none of these relationships is statistically significant).

Figure 3 presents the relationship between cohort size and class size, where again each data point is at the year–degree level. The correlation is positive, but insignificant, for all subjects: students in larger cohorts are more likely to be taught in large classes.

The large variation in teaching arrangements across subjects and universities implies that students have markedly different experiences. The provision of such information to current and potential students would increase

TABLE 2

Contact hours per week by group size

(a) Lectures

	Lecture size						*Total*
	>200	151–200	101–150	51–100	26–50	<26	
Physics	1.29	0.75	1.49	2.99	1.62	0.81	8.95
History	0.63	0.75	0.96	2.25	1.14	1.00	6.73
Economics	3.02	1.07	1.41	1.49	0.69	0.17	7.85

(b) Classes

	Class size					*Total*
	>40	26–40	16–25	6–15	1–5	
Physics	0.69	0.35	0.59	0.23	0.20	2.05
History	0.10	0.17	1.22	1.05	0.09	2.64
Economics	0.18	0.37	1.32	0.93	0.08	2.89

(c) Practicals

	Practical size					*Total*
	>40	26–40	16–25	6–15	1–5	
Physics	0.75	0.40	0.42	0.94	0.30	2.82
History	0.01	0.01	0.02	0.02	0.03	0.09
Economics	0.10	0.05	0.04	0.05	0.02	0.26

FIGURE 2

Contact hours and class size

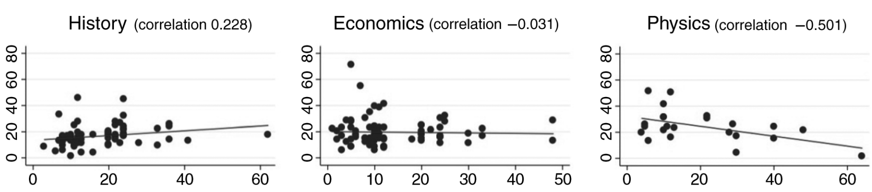
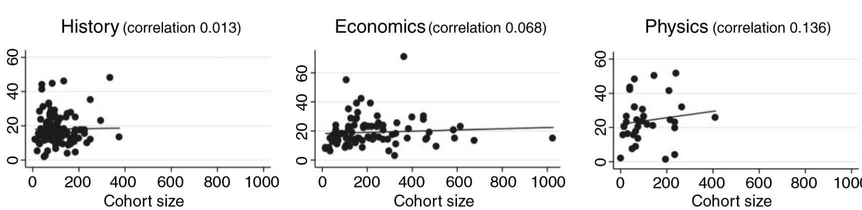


FIGURE 3

Cohort size and class size



transparency and choice. The following section develops a measure of teaching intensity that makes meaningful comparisons across universities.

VI. Total equivalent adjusted contact hours (TEACH)

One key objection to direct comparison of contact hours as a measure of ‘overall teaching’ is that teaching may consist of different activities and/or be in different-sized classes. We overcome this problem by defining an ‘equivalent adjusted contact hour’ (EACH) as an hour of time with a teacher weighted by $1/n$, where n is the number of students present in the session. For example, 1 EACH could be 1 hour of one-to-one teaching or equivalently it could be 10 hours of a group of 10 students in a tutorial.

For a student studying a particular course at a particular university, summing these adjusted contact hours provides us with the ‘total equivalent adjusted contact hours’ (TEACH):

$$(1) \qquad TEACH = w_1 q_1 + w_2 q_2 + \cdots + w_n q_n + \cdots,$$

where q_1 is the number of hours in a one-to-one class, …, q_{10} is the number of hours in a class with 10 students, etc. w_n is the weight ($1/n$) given to classes of size n. TEACH is therefore a measure of the labour time per student.

To clarify, an EACH is not a measure of welfare and makes no assumptions about the benefits of different class sizes. Our weighting of $w_n = 1/n$ is in terms of the resource cost, with the weighting corresponding to the market trade-off between contact hours and intensity. Other weightings would impose homogeneous preferences on all students that do not correspond to this trade-off. Two obvious examples: using a weighting of $w_n = 1/\sqrt{n}$ would suggest a preference for contact hours over intensity, whereas using a weighting of $w_n = 1/n^2$ would suggest a preference for intensity over contact hours.

If all students have the same preferences and are fully informed, then class size should converge. If students have different preferences, universities will differentiate themselves in terms of provision.

Figure 4 shows the distribution of TEACH across subjects. There is huge variation in TEACH both between and within subjects.[28] The ratios of the maximum to minimum TEACH in economics, history and physics are respectively 21.6, 6.4 and 25.8. Physics students receive the largest mean TEACH (74.6 over a three-year degree), approximately 2.3 times that of history (32.6 TEACH) and 2.9 times that of economics (26.1 TEACH).[29]

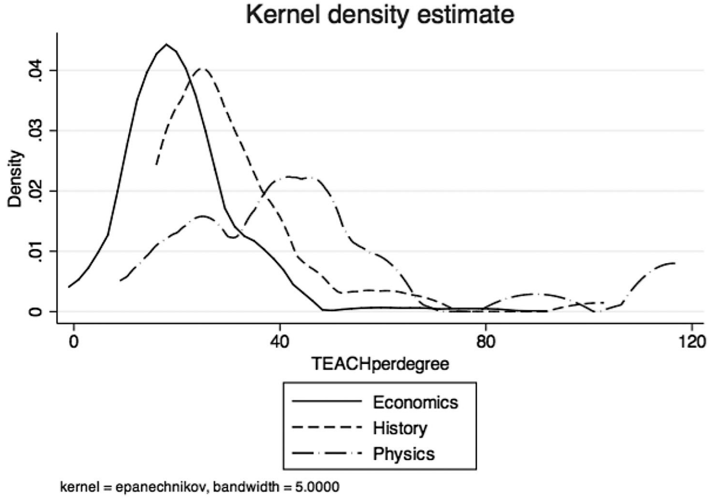
Although measures of TEACH are useful in ranking resources provided by different universities, how these data are presented to prospective students

[28]This is even after omitting the five universities that offered the highest TEACH. Removing these ‘outliers’ makes the diagram clearer.

[29]The median TEACH per degree for physics, history and economics are 46.2, 31.5 and 20.4 respectively.

FIGURE 4

Distribution of TEACH by subject



Note: The five universities that offered the highest TEACH have been omitted.

also matters. No student receives all their teaching in the form of one-to-one classes, so an alternative way of presenting TEACH would be a mapping onto a more typical student experience. For example, 10 hours of lectures (100 students) and 6 hours of seminars (15 students) per week, over three years of 24 teaching weeks, would result in 36 TEACH.

This measure of teaching helps to clarify some of the relevant trade-offs. For example, the TEACH constraint is non-linear in class size: larger class sizes result in progressively smaller increases in contact hours. Larger lectures require fewer EACHs per student and this releases resources that can be used to reduce class size. However, this effect is only present for small cohorts. Increasing an already large lecture brings progressively smaller benefits in terms of class size reduction. If large lectures result in reduced quality, students are justified in resisting larger cohorts or insisting on the duplication of lectures.

Although EACHs allow us to compare different bundles of teaching, they say nothing about which bundles might be preferred by students. For example, students might prefer an extra contact hour to a reduction in class size. In earlier work,[30] we consider how students react to the ‘intensity’ of the teaching they receive. In that model, we show that the demand for teaching will depend on the degree of complementarity between teaching intensity and independent study.

[30]Huxley and Peacey, 2014.

VII. Empirical analysis

In this section, we address two questions. First, what is the nature of the relationship between teaching and the key characteristics of both the degree programme and the university? Answering this question will show whether the provision of information on teaching intensity to students is likely to provide additional information over and above other characteristics such as Russell Group status and research intensity. Second, can TEACH be used to explain student satisfaction results? A weak relationship would suggest either that TEACH fails to capture what students care about or that students cannot benchmark the teaching they receive.[31]

In the first model, we use a standard linear regression to explain TEACH per degree:

$$(2) \qquad \ln TEACH_{ijk} = \alpha + \beta X_{ijk} + \delta Z_j + \epsilon_{ijk}.$$

We interpret equation 2 as a teaching production function, where i represents degree, j represents university and k represents degree year. X_{ijk} is a vector of degree–year-specific variables such as cohort size and proportion of classes taught by those on hourly-paid contracts, Z_j is a vector of university-specific variables such as endowment[32] and location,[33] and ϵ_{ijk} is an error term, clustered at the institution level. A log-linear model is chosen because we would expect diminishing marginal effects of degree- and university-specific variables on TEACH. We also include a dummy variable for the Russell Group,[34] a collection of prestigious public research universities.

In the second set of models, we again use simple log-linear regressions, this time including the data we have collected as explanatory variables, to explain a variety of questions in the NSS:

$$(3) \qquad \ln NSSQUESTION_{ijk} = A + BX_{ij} + CY_j + DW_k + e_{ijk},$$

where i represents degree, j represents university and k represents the department (it captures all courses offered by the economics department, for example). X_{ij} is the same vector as before, with the addition of *TEACH* and other characteristics that might affect NSS scores (for example, use of hourly-paid staff, the proportion of classes with at most nine students and the proportion of part-time students), Y_j is the university-specific variables

[31] Brown et al., 2015.

[32] We used endowments as reported by universities in official audited financial statements in 2013. Where data were not available for 2013, we used the closest available financial year.

[33] We used 12 locations: the nine geographical regions of England, plus Scotland, Wales and Northern Ireland.

[34] Eleven of the universities in our data set are members of the Russell Group. We tried controlling for membership of other groups, such as the University Alliance, but there were problems of multicollinearity.

included in Z_j, and W_k is the department-level variables (for example, proportion of female students and proportion of white students) collected from the Heidi database.[35] The error term e_{ijk} is again clustered at the institution level. As well as the controls used in equation 2, this model includes degree-level variables such as the proportion of students who obtained first-class honours. All of this information was included separately to allow for independent effects on NSS scores, but this means that the magnitude of the TEACH variable may not be identified. As there may be unobserved factors that affect student satisfaction, we cannot claim causality.

We are interested in the regression because it shows what, if anything, we can learn about NSS scores from currently available data. Whilst the NSS is important because it will be one of the metrics used to assess teaching in the TEF, it should be noted that it is an imperfect measure of student satisfaction[36] and is unreliable as a measure of education quality.[37]

1. What explains the large variation in TEACH?

In this section, we seek to explain what determines how much teaching universities offer. A Chow test revealed that the coefficients on some of the explanatory variables, such as Russell Group membership, were significantly different for physics compared with economics and history. Thus, for this regression, we split the data: Table 3 presents the results for physics separately from those for economics and history. The table shows the results of equation 2, with only the location variables not shown.

Membership of the Russell Group has a dramatically different effect on TEACH per degree for physics compared with economics and history. The first column of Table 3 suggests that Russell Group universities provide 60 per cent more TEACH per degree to physics students than non-Russell-Group universities. This is in contrast to the results in the second column, which suggest that Russell Group universities provide 38 per cent less TEACH per degree to economics and history students than non-Russell-Group universities. Thus, differences between subjects are largest amongst the Russell Group.

We also find that TEACH is increasing in research strength for all subjects.[38] However, conditional on Russell Group membership, there is a weak positive correlation between endowment (measured in £ million) and TEACH per degree for economics and history, suggesting that wealthy universities can 'afford' better arts, humanities and social science teaching.

[35] Heidi is a database run by HESA and is a source of quantitative data on higher education providers.
[36] Gibbons, Neumayer and Perkins, 2015.
[37] Brown et al., 2015.
[38] Research strength is calculated using the 2008 Research Assessment Exercise data by multiplying the research score by the number of staff who submitted. It has a mean of 41 and a standard deviation of 54.

TABLE 3

TEACH per degree

	Physics	*Economics and history*
Hourly-paid staff used for classes	–0.382	0.147
	(0.494)	(0.162)
Cohort size	–0.00168	–0.00134**
	(0.00112)	(0.000618)
Russell Group	0.599	–0.384
	(0.537)	(0.243)
Fee (£'000)	0.0689	–0.361
	(0.203)	(0.236)
Endowment (£m)	–0.00943**	0.00294
	(0.00396)	(0.00260)
Research strength	0.00874***	0.00520
	(0.00224)	(0.00392)
No. of observations	75	207
R^2	0.739	0.359

Note: Location is included in the regressions but not displayed. Standard errors are given in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

The coefficient on fees is close to zero and insignificant for physics, but negative and close to significant for economics and history students. Students who pay higher tuition fees do not receive more tuition. Prima facie, these findings are hard to reconcile with the rationale for variable tuition fees put forward in the 2003 and 2011 White Papers.[39] The variation in TEACH between subjects might be explained in terms of unobserved characteristics such as teacher quality, but universities need to account for these differences.

The regression tries and mostly fails to explain the determinants of TEACH (using variables that are not usually observed by students as well as those that are). The finding that TEACH cannot be inferred from the characteristics that students can observe suggests that TEACH might provide valuable information to students, above that already observed through subject and university characteristics.

2. Does TEACH affect student satisfaction?

This subsection shows the relationship between self-reported student satisfaction and our measure of teaching intensity, as well as other potential determinants of student satisfaction. This allows us to see whether teaching

[39]Department for Education and Skills, 2003; Department for Business, Innovation and Skills, 2011.

intensity is related to student satisfaction under the current arrangements, where a lack of transparency makes comparison by students across institutions hard.

Table 4 presents the results of the NSS regression model. The NSS contains 22 questions, split into five categories, and an overall satisfaction score. The categories are: teaching; assessment and feedback; academic support; organisation, management and resources; and personal development. We provide summary statistics for these categories, split by subject, in Appendix D and we run regressions with the averages of these as the dependent variables in Appendix E.[40]

The dependent variable in Table 4 is the proportion of students who respond 'agree' or 'definitely agree' to the statement 'Overall, I am satisfied with the quality of my course'. Although this question asks explicitly about the course quality, it is likely that answers are confounded by other dimensions of the student experience.

Economics and history students are less satisfied than physics students (with overall satisfaction down 8 per cent and 6 per cent respectively, although the result is not statistically significant). The coefficient on fees suggests that a £1,000 increase in fees is associated with a reduction in overall satisfaction of 2.4 per cent, although it is insignificant. The coefficient on endowment is negative but close to zero.

Many of these results are mirrored when NSS scores are inspected by category (in Appendix E) – especially the relative dissatisfaction of economics and history students. However, in most of these categories, TEACH per degree was slightly positively correlated with satisfaction – although the coefficient was always insignificant. Other than teaching satisfaction, little else can be explained by the models, supporting the criticisms of the NSS discussed in Section II.

These findings raise questions about the use of NSS scores as a proxy for learning gain. Learning gain in education is difficult to measure, and especially so in higher education because university students do not sit standardised tests. The results of standardised tests in primary and secondary school mean that data on pupil outcomes can be observed by parents, teachers and researchers.[41] In higher education, this information is not available to any of the relevant stakeholders. We would expect at least some of the variables in our regression to influence outcomes; the fact that they do not suggests that either the input or the output measures fail to capture what matters. We believe that TEACH will provide researchers with some of the information required to better understand the determinants of learning gain in higher education. Hoxby (2014) uses the frequency of instructor–student interactions as a measure of quality in higher education. Furthermore, the unfunded expansion that took place between 1980

[40]Both appendices are available online.
[41]Arum and Roksa, 2011.

TABLE 4

Overall satisfaction

	Overall satisfaction
TEACH per degree	–0.0000698
	(0.0000604)
Proportion of classes that are 'small'	0.0206
	(0.0528)
Cohort size	0.0000470
	(0.0000477)
History	–0.0615
	(0.0693)
Economics	–0.0791
	(0.0536)
Hourly-paid staff used for classes	0.00464
	(0.0184)
Hourly-paid staff used for lectures	–0.0204
	(0.0270)
Hourly-paid staff used for practicals	–0.0745
	(0.0511)
Fee (£'000)	–0.0239
	(0.0319)
Russell Group	–0.0249
	(0.0197)
Endowment (£m)	–0.000468***
	(0.000162)
Research strength	0.000187
	(0.000174)
Proportion of students with first-class honours	0.000436
	(0.000672)
Proportion of part-time students	0.169
	(0.106)
Proportion of female students	0.220
	(0.196)
Proportion of disabled students	–0.388
	(0.351)
Proportion of white students	0.118
	(0.105)
No. of observations	219
R^2	0.422

Note: Location is included in the regression but not displayed. Standard errors are given in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

and 2000, and its implications for the funding per student, staff–student ratios and class size, formed the core of the case for the introduction of tuition fees.[42] Therefore if TEACH is unrelated to measures of learning gain, the standard arguments for increasing tuition fees start to unravel. The implication for cost (and therefore the fee) must be understood by the sector.

3. Sensitivity analysis[43]

In this subsection, we present the results of three robustness checks. First, we experimented with adjusting the weighting used for calculating EACHs (see Section VI). When using weightings of $w_n = 1/\sqrt{n}$ or $w_n = 1/n^2$, the magnitude of the coefficients naturally altered, but the signs and significance were unchanged. Second, we ran the NSS regressions using only the proportion of students who answered 'definitely agree' to each question. The overall picture was unchanged, with only some coefficients increasing in magnitude and a few insignificant coefficients changing signs. Third, we tried including practicals in our calculations of class size. Again, the overall picture was unchanged, except for the coefficient on 'Proportion of classes that are small' for economics and history. This is because practicals are, on average, much larger than classes in economics and history, whilst they are similar sizes in physics.

Finally, we looked at the variation in TEACH provided by each component part: lectures, classes and practicals. There is most variation in the TEACH provided by lectures for physics. For classes, economics and history display the greatest differences in TEACH offered.

VIII. Conclusion

Our data reveal large differences in teaching intensity across higher education in the UK. We have shown that the usual explanations for this variation are unsatisfactory. For example, it is not true that variation in contact hours can be explained by variation in class size. These differences exist both within disciplines and between disciplines, and are so large it is hard to see how they can be explained by offsetting differences in other dimensions of the student experience.

In terms of our TEACH metric, some students receive much better value for money than others. For example, economics students in the top decile of the TEACH distribution receive almost five times as much teaching as students in the bottom decile. We also find large differences between the three subjects:

[42] Greenaway and Haynes, 2003.
[43] Full results of the analysis in this subsection are available from the authors upon request.

students studying economics and history receive on average less than half the teaching received by physics students.

We know that there are differences in private returns across subjects,[44] which may result from signalling or human capital.[45] The optimal amount of TEACH per degree will depend on which hypothesis is subscribed to.[46] If, for whichever reason, differences in TEACH across subjects are required, this should be reflected in the fee.

If the social return to human capital investments in STEM subjects exceeds the private return, a compensating subsidy can be justified. Only if more teaching is required to generate the externality will the subsidy come in the form of an increase to TEACH. If instead the need is for more graduates, the subsidy should be designed to increase the supply of students through lower fees.

Some subjects receive a subsidy from the Treasury depending on their price band classification by the Higher Education Funding Council for England (HEFCE). The standard explanation for these differences is that STEM subjects have higher fixed costs (laboratory equipment, etc.) than arts, humanities and social science.[47,48] It has always been the case that students studying STEM subjects have more contact hours but, in the past, this was offset by larger classes.[49] There will also be differences in wage cost and quality that we do not measure, which might account for some of the observed differences in TEACH.

We find that students studying economics and history receive less teaching than students studying physics, and this discrepancy is much larger than can be explained by the additional tax funding for STEM subjects provided by the Treasury.[50] It follows that either the fee or the subsidy has been set incorrectly: arts, humanities and social science students should receive more teaching or STEM students should receive a larger subsidy.

The observed variation in weighted units of teaching can be partially explained by the type of university: at Russell Group universities, students studying physics are advantaged whilst students studying economics or history are penalised. However, little of the variation in teaching can be explained

[44]Britton, Shephard and Vignoles, 2015.

[45]Weiss, 1995.

[46]The case for variable fees has usually been made in terms of the human capital hypothesis (Barr, 2004).

[47]Higher Education Funding Council for England, 2005.

[48]The latter are often referred to as library-based disciplines. See, for example, this plea for economics to be reclassified: http://www.res.org.uk/view/art3Apr04Features.html.

[49]Robbins, 1963.

[50]Physics is a band B subject and in 2013–14 received £1,500 per year extra from HEFCE on top of the £9,000 fee paid by students, while economics and history are both band D subjects and receive zero teaching subsidy (Higher Education Funding Council for England, 2014). Thus the unit of resource for physics (£9,000+£1,500) is 116.7 per cent of that for economics or history. We find mean TEACH for physics is 230 per cent of that for history and 290 per cent of that for economics.

by factors such as the fee, research strength or endowment: accident and history appear to determine the outcome. We found that the National Student Survey results cannot be satisfactorily explained by either the existing data (for example, endowment or research strength) or our new administrative data (for example, TEACH or cohort size). However, the NSS regressions do suggest that physics students appear to be more satisfied than students studying economics or history. One possible explanation for this finding is that physics students perceive that, relative to students studying other subjects, they get better value for money.[51]

In Salop and Stiglitz (1977), the welfare properties of the equilibrium depend on the fraction of informed consumers. We believe the enormous variation in teaching intensity found in our data strongly suggests that price signals are weak in the market for teaching, and this raises important issues for the design of policy. This does not imply that market forces are absent in higher education. In those dimensions that can be observed, universities and individual academics operate in a highly competitive environment. Students also compete for places at the most prestigious universities, and universities compete for the highest-ability students. Unfortunately, little of this competition is in terms of the teaching offered by universities.

In an announcement about the Teaching Excellence Framework, Universities Minister Jo Johnson stated that 'universities must get used to providing clearer information about how many hours students will spend in lectures, seminars and tutorials, and who will deliver the teaching'.[52] Our framework allows a student choosing between courses to focus on three questions: the units of teaching on offer (i.e. TEACH); how these resources are allocated across different learning activities (i.e. lectures, classes and practicals); and the teaching intensity (i.e. class size) of each of these learning activities. Presenting data in this way permits students to choose the university that offers their preferred trade-off between contact hours and class size. Universities should remain free to deliver teaching in any way they choose, but this autonomy must be reconciled with accountability to students.

In this paper, we have shown that administrative data exist that can be used to provide information, at a course level, on an important dimension of quality in higher education. In our experience, many universities already store this information centrally, with many others holding it at the faculty or department level. Even universities offering a modular degree structure were able to provide the data. As a point of historical record, the Robbins Report committee used Hollerith machines to analyse the data it collected. Given all the advances in data recording, storage and analysis that have taken place since

[51] Brown et al., 2015.
[52] Johnson, 2015.

1963, it is both feasible and reasonable to expect universities to collate and report meaningful data on inputs offered to prospective students.

Any metric included in the Teaching Excellence Framework should be precisely measurable, be hard to manipulate and actually matter. TEACH meets these objectives. Unlike measures of contact hours or class size, there is no resource-free way to game the TEACH metric. The only way to increase TEACH is to increase teaching – increasing contact hours funded by increased class sizes will not alter TEACH. Precisely how TEACH should fit into the TEF is beyond the scope of this paper. We restrict ourselves to two observations. First, in contrast to the Research Excellence Framework, which is able to make use of the pre-existing (and widely accepted) metric that is the peer review assessment of research, the TEF has to build a metric from scratch. Second, we have reservations about TEACH simply being used as an input in any TEF score. The main contribution of our metric is to enable prospective students to make better-informed decisions by providing information on the teaching arrangements on undergraduate degree programmes they are considering.

We are in no doubt that there are many dimensions of the student experience other than teaching that contribute to the quality of higher education. For example, libraries, sports facilities, accommodation and extracurricular activities are all important. Relative to teaching, these characteristics are readily observed and, as such, universities already compete to offer them in increasingly high quality. The critical allocation problem in higher education is between teaching and research. Our metric is unique in its focus on this margin and we believe, if adopted, it will help rebalance this relationship.

Finally, it should be obvious that additional TEACH may not be beneficial for all students – for example, if the cost of extra contact is lower teacher quality. Even when quality is held constant, some students may be better off working on their own if the time cost of attending extra classes is self-study. The case for providing students with information on TEACH depends upon an understanding of the link between TEACH and outcomes. This will require a better understanding of the relationship between inputs used in higher education and the outcomes they generate. The necessary research can only be undertaken if universities are transparent about the teaching they deliver.

Supporting information

Additional supporting information may be found in the online version of this paper on the publisher's website:

- Appendix A. Comparing Robbins and HEPI with our data
- Appendix B. The FoI request
- Appendix C. Summary of sample
- Appendix D. NSS summary statistics

- Appendix E. NSS regressions by category
- Appendix F. Lists of universities
- Appendix G. The modified FoI request

References

Allgood, S., Walstad, W. B. and Siegfried, J. J. (2015), 'Research on teaching economics to undergraduates', *Journal of Economic Literature*, vol. 53, pp. 285–325.

Arias, J. and Walker, D. M. (2004), 'Additional evidence on the relationship between class size and student performance', *Journal of Economic Education*, vol. 35, pp. 311–29.

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

Badri, M. A., Abdulla, M., Kamali, M. A. and Dodeen, H. (2006), 'Identifying potential biasing variables in student evaluation of teaching in a newly accredited business program in the UAE', *International Journal of Educational Management*, vol. 20, pp. 43–59.

Bandiera, O., Larcinese, V. and Rasul, I. (2010), 'Heterogeneous class size effects: new evidence from a panel of university students', *Economic Journal*, vol. 120, pp. 1365–98.

Barr, N. (2004), 'Higher education funding', *Oxford Review of Economic Policy*, vol. 20, pp. 264–83.

Becker, W. E. and Powers, J. R. (2001), 'Student performance, attrition, and class size given missing student data', *Economics of Education Review*, vol. 20, pp. 377–88.

Britton, J., Shephard, N. and Vignoles, A. (2015), 'Comparing sample survey measures of English earnings of graduates with administrative data during the Great Recession', Institute for Fiscal Studies (IFS), Working Paper no. W15/28.

Brown, G. D., Wood, A. M., Ogden, R. S. and Maltby, J. (2015), 'Do student evaluations of university reflect inaccurate beliefs or actual experience? A relative rank model', *Journal of Behavioral Decision Making*, vol. 28, pp. 14–26.

Cheng, J. H. and Marsh, H. W. (2010), 'National Student Survey: are differences between universities and courses reliable and meaningful?', *Oxford Review of Education*, vol. 36, pp. 693–712.

De Fraja, G. and Valbonesi, P. (2012), 'The design of the university system', *Journal of Public Economics*, vol. 96, pp. 317–30.

Department for Business, Innovation and Skills (2011), *Higher Education: Students at the Heart of the System*, Cm 8122, Stationery Office.

— (2016), *Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice*, Cm 9258, Stationery Office.

Department for Education (2016), *Teaching Excellence Framework: Year Two and Beyond*, Stationery Office.

Department for Education and Skills (2003), *The Future of Higher Education*, Cm 5735, Stationery Office.

Gibbons, S., Neumayer, E. and Perkins, R. (2015), 'Student satisfaction, league tables and university applications: evidence from Britain', *Economics of Education Review*, vol. 48, pp. 148–64.

Gibbs, G. (2010), *Dimensions of Quality*, York: Higher Education Academy.

— (2012), *Implications of Dimensions of Quality in a Market Environment*, York: Higher Education Academy.

Greenaway, D. and Haynes, M. (2003), 'Funding higher education in the UK: the role of fees and loans', *Economic Journal*, vol. 113, pp. F150–66.

Hale, E. and Tattersall, A. (1964), *University Teaching Methods: A Report of the Committee on University Teaching Methods*.
Higher Education Funding Council for England (2005), *Review of the Teaching Funding Method*.
— (2014), *Guide to Funding and Student Number Controls 2013–14 and 2014–15: How HEFCE Allocates Its Funds and Controls Student Numbers*.
Higher Education Policy Institute (2016), *HEPI Student Academic Experience Survey*.
Hillman, N. (2015), 'Data-led policy: what should the 2015 HEPI / HEA Academic Experience Survey mean for policy?', Higher Education Policy Institute (HEPI) blog.
Hoxby, C. M. (2014), 'The economics of online postsecondary education: MOOCs, nonselective education, and highly selective education', National Bureau of Economic Research (NBER), Working Paper no. 19816.
Huxley, G. and Peacey, M. W. (2014), 'Modelling learning parameters', University of Bristol, Department of Economics, Working Paper no. 13/322.
Huybers, T., Louviere, J. and Islam, T. (2015), 'What determines student satisfaction with university subjects? A choice-based approach', *Journal of Choice Modelling*, vol. 17, pp. 52–65.
Johnson, J. (2015), 'Teaching at the heart of the system', speech, July.
Kokkelenberg, E. C., Dillon, M. and Christy, S. M. (2008), 'The effects of class size on student grades at a public university', *Economics of Education Review*, vol. 27, pp. 221–33.
Krueger, A. B. (2003), 'Economic considerations and class size', *Economic Journal*, vol. 113, pp. F34–63.
Lenton, P. (2015), 'Determining student satisfaction: an economic analysis of the National Student Survey', *Economics of Education Review*, vol. 47, pp. 118–27.
Martins, P. S. and Walker, I. (2006), 'Student achievement and university classes: effects of attendance, size, peers, and teachers', Institute for the Study of Labor (IZA), Discussion Paper no. 2490.
Monks, J. and Schmidt, R. M. (2011), 'The impact of class size on outcomes in higher education', *BE Journal of Economic Analysis & Policy*, vol. 11, issue 1, https://doi.org/10.2202/1935-1682.2803.
Richardson, J. T., Slater, J. B. and Wilson, J. (2007), 'The National Student Survey: development, findings and implications', *Studies in Higher Education*, vol. 32, pp. 557–80.
Robbins, L. R. B. (1963), *Higher Education: Report of the Committee Appointed by the Prime Minister under the Chairmanship of Lord Robbins, 1961–63*, no. 2, HM Stationery Office.
Salop, S. and Stiglitz, J. (1977), 'Bargains and ripoffs: a model of monopolistically competitive price dispersion', *Review of Economic Studies*, vol. 44, pp. 493–510.
Sapelli, C. and Illanes, G. (2016), 'Class size and teacher effects in higher education', *Economics of Education Review*, vol. 52, pp. 19–28.
Schanzenbach, D. (2010), 'The economics of class size', in D. J. Brewer and P. J. McEwan (eds), *Economics of Education*, Amsterdam: Elsevier.
Weiss, A. (1995), 'Human capital vs. signalling explanations of wages', *Journal of Economic Perspectives*, vol. 9, no. 4, pp. 133–54.