Class Size at University*

Gervas Huxley[†], Jennifer Mayo[‡],
Mike W. Peacey[§†] and Maddy Richardson[†]

May 30, 2017

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 is currently available. We define a measure of teaching that weights contact hours by their intensity and collect a new dataset that allows comparison of teaching across universities and three departments.

No two universities offer identical teaching. There is large variation in contact hours and even larger variation in teaching intensity - both across universities and departments. We combine our data with existing data to investigate the relationship that teaching has with university and student characteristics. We find that how much teaching students receive is uncorrelated with tuition fee; that teach- in has little predictive power in explaining student satisfaction; and that Physics students consistently receive more teaching than either Economics or History students.

Keywords: Class Size, Student Choice, Teaching Excellence Framework (TEF), Teaching Intensity.

JEL: I23, I28

∗ This project benefited from funding from the Centre for Markets and Public Organisation (CMPO). We are grateful for comments from seminar participants at New College of the Humanities, ISER (Essex), LEER (Leuven) and AEDE (Madrid). We are also grateful to: Gill Clarke (for bringing to our attention the appendices to the Robbins report), and for comments from Jack Britton, Olga Gdula,

Ellen Greaves, Nick Hillman, Maria Mullens-Burgess, Helen Simpson, David Willetts, and David Winter.

† University of Bristol

‡ University of Michigan

§ New College of the Humanities (e-mail: mike.peacey@nchlondon.ac.uk)

Policy points

- In contrast to many dimensions of quality (e.g. 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. In Physics students receive 2.3 times more than in History and 2.9 times more than Economics. The ratio of maximum to minimum teaching provided across universities is 21.6, 6.4, and 25.8 for Economics, History, and Physics respectively.
- We argue that this measure can complement existing metrics to increase transparency and improve student choice.

'A competitive and dynamic university sector requires students who actively and regularly challenge universities to provide teaching excellence and value for money."[1]

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 (De Fraja and Valbonesi (2012)). This means universities must compete in terms of characteristics that students care about. To achieve this student must 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 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

[1] BIS (2016).

lecture, tutorial, seminar, or laboratory. Second, even if agreement can be reached on what is meant by these terms, universities offer very different bundles and this 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, personal characteristics, and 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 Higher Education Policy Institute is completed by approximately 15,000 full-time undergraduates (HEPI (2016)). 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 (see Section III.). 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), 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: *'No actors in the system are primarily interested in undergraduate student academic growth'* (Arum and Roksa (2011)). 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 (BIS (2016)). 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 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 and introduce our teaching intensity metric. In Section V. we present summary statistics of our data to describe the current state of teaching in HE. In Section VI. we explain the problems that arise in interpreting these data, and introduce our proposed metric. In Section VII. we combine our data with existing publicly available data to analyze 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’ (Gibbs (2010)) 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[2].

There is a small literature on class size in higher education (e.g. Bandiera et al. (2010), Martins and Walker (2006), Monks and Schmidt (2011), Sapelli and Illanes (2016)). 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 is 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.

[2] 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.

Bandiera et al. (2010) use administrative data on individual students from a leading UK university who are enrolled in one-year MSc programme 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 1-19 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 reduction in class size.

In the context of teaching undergraduate economics Becker and Powers (2001), Arias and Walker (2004), and Kokkelenberg et al. (2008) all find that large classes are associated with negative outcomes such as higher dropout rates for weak students, exam performance, progression to advanced level courses, and student evaluations (see Allgood et al. (2015) for a survey on teaching undergraduate economics). 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 et al. (2015) observe: *'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 (for an econometric analysis of the UK National Student Survey (NSS) see Lenton (2015)). In the UK, it has been proposed that a university's benchmarked score in the NSS will partly determine whether it is permitted to raise its undergraduate fee (DfE (2016)). The motivation for introducing the NSS was to inform and influence student choice (Richardson et al. (2007)). However, Gibbons et al. (2013) find that the additional information provided by the NSS *'has only a small impact on the choices of students'*. They find that improvements which move the NSS score from the median to 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 (e.g. age, gender, previous schooling, etc. - see Badri et al. (2006) and Huybers et al. (2015)). 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 may be influenced by factors other than objective quality of educational experience.'*

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 which include studies of the student experience. 4224 undergraduate students (1 in 22 of the student population at the time) completed detailed time sheets setting out how they used their time in February 1962[3]. 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[4]. 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 (Appendix III p.73). In 1963 students complained about the quality of their education, but the majority of such complaints were focused on

---

[3] This data is presented in the Hale Report (Hale and Tattersall (1964)) and in Appendix II and III of the Robbins Report (Robbins 1963).

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

demands for 'Oxbridge style' tutorials. In Appendix 2 we investigate how contact hours and class size have changed since 1963, 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: in the humanities 70 per cent and in social science 65 per cent of students 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 (Appendix II B) finds that PhD students did some of the teaching - but far less than today. PhD students were used in science subjects to take practicals (but not, by and large, to teach); in arts, humanities, and social science, hardly at all.

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 subject: for example, just over 17 hours for physical sciences vs less than nine 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 2 we compare these results to the data we have collected.

IV. Administrative data collection

The increase in 2012-13 of tuition fees to £9,000pa 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 FoI Act.

The FoI Act created a right to request certain information held by public authorities, subject to certain exemptions[5]. For the purpose of the FoI Act universities are generally regarded as public authorities[6].

[5] 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 parties' commercial interests (subject to a public-interest test).

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

Our FoI request (Appendix 1) 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 -whilst keeping the request to a size that could be met by the university 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 provided us with contact hours, broken down into lectures and classes[7]. By combining this with module enrollment we were able to calculate (amongst other things) mean class size.

Of the 99 universities we contacted, four were either 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[8]. Our final dataset contains 67 universities: 59 of these provided us with the data

[7] We defined a 'class' as any teaching where students are divided in groups smaller than the number enrolled in lectures - but which is not classified as a 'practical' (see Appendix 1). In practice, universities usually call the groups either tutorials or seminars.

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

following the initial request and eight agreed to the data requested following our modified FoI request (Online Appendix 2)[9].

In Appendix 3 we compare the characteristics of the universities who did and did not provide data we asked for[10]. 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

[9] 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.

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

unobservable to students. Table 1 shows the mean and standard deviation for the main variables in our dataset[11].

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[12]. There is large variation: lectures given to the smallest of these cohorts will feel more like a class.

Some modules deliver all their teaching via lectures and practicals. In Physics 37 per cent of modules deliver their teaching via classes compared to 84 per cent of Economics modules. Modules that offer no classes are excluded from our calculation of average class size[13]. 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 and this can be seen in Figure 1. The variable *proportion of classes that are small* is the proportion of teaching provided in classes of fewer than nine students, and is calculated as a proportion of classroom teaching. History students are taught in classes of fewer than nine the most often.

[11] 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.

[12] 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.

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

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) compared to History (10 per cent of classes). The other three variables report the total hours of lectures, classes and practicals for each year group.

Table 2 presents the data on contact hours, split by (a) lectures, (b) classes, and (c) practicals. These group sizes are inevitably arbitrary and we have chosen ones which may help the reader gain a picture of the distribution. These tables illustrate how misleading it is to give students information on contact hours without providing any, or uninformative, 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.

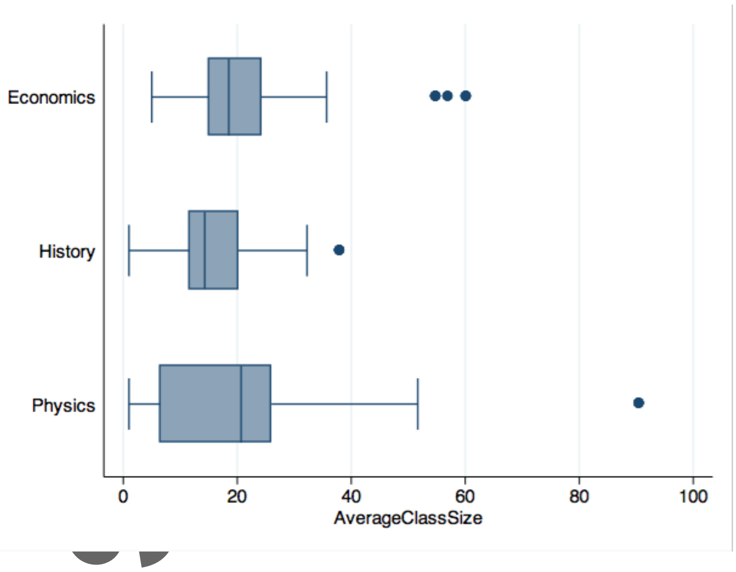

FIGURE 1: Average class size by subject

TABLE 1: Summary statistics

	Economics (SD)	History (SD)	Physics (SD)
Cohort size (/year)	215.83 (170.34)	104.11 (67.49)	110.46 (79.06)
Proportion of modules that	.84 (.24)	.53 (.39)	.37 (.37)
Average lecture size	111.93 (115.19)	41.19 (41.66)	76.52 (73.88)
Average class size	19.89 (12.04)	15.92 (8.38)	21.50 (12.85)
Average practical size	66.48 (103.02)	36.24 (42.73)	25.09 (21.76)
Total hours of lectures	169.77 (104.61)	148.07 (73.61)	201.28 (119.76)
Total hours of classes (/year)	68.91 (45.59)	61.39 (68.34)	46.17 (76.47)
Total hours of practicals	6.38 (12.52)	2.15 (8.59)	68.38 (70.67)
Proportion of classes that are	.07 (.15)	.11 (.15)	.05 (.06)
Proportion of classes taught	.16 (.20)	.10 (.12)	.15 (.29)
Observations	95	133	79

TABLE 2

Contact hours grouped by class size

(a) Lectures

	>200	151 to 200	101 to 150	51 to 100	26 to 50	<26	TOTAL
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

	>40	26 to 40	16 to 25	6 to 15	1 to 5	TOTAL
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

	>40	26 to 40	16 to 25	6 to 15	1 to 5	TOTAL
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

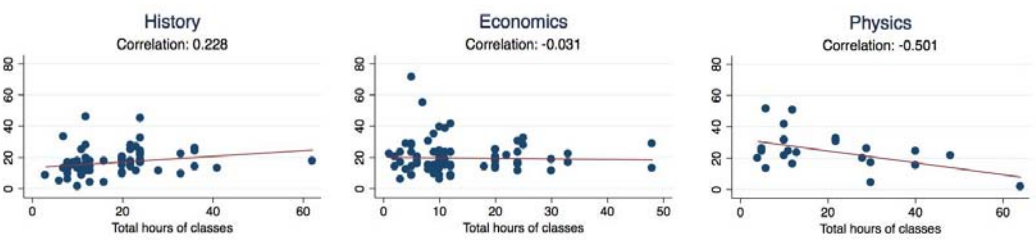

FIGURE 2

Class contact hours and class size

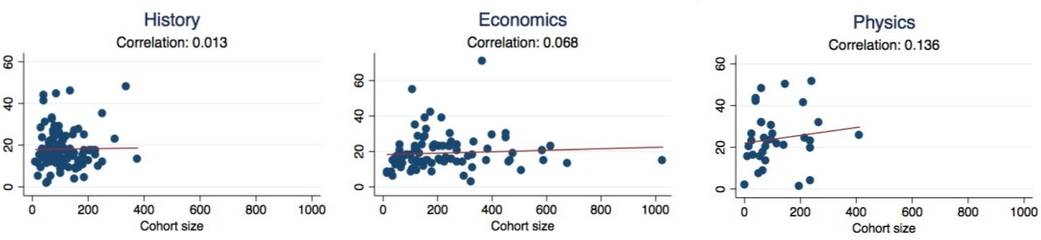

FIGURE 3

Cohort size and class size

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 are 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 transparency and choice. The following section develops a measure of teaching intensity that makes meaningful comparable 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 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 $\frac{1}{n}$, where n is the number of students present in the session. For example, one 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):

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

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 $\left(\frac{1}{n}\right)$ given to classes of size n. TEACH is therefore a measure of the labor 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 $w_n = \frac{1}{n}$ 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 homogenous preferences on all students that do not correspond to this trade-off. Two obvious examples: using a weighting of $w_n = \sqrt{n}$ would suggest a preference for contact hours over intensity, whereas using a weighting of $w_n = n^2$ would suggest a preference for intensity over contact hours.

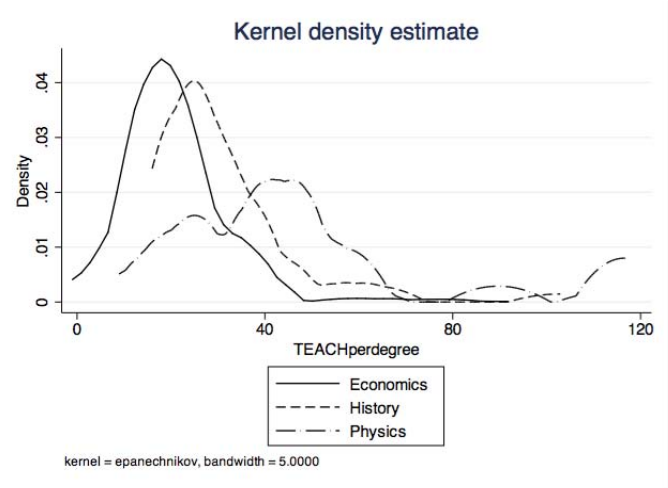

FIGURE 4

Distribution of TEACH by subject

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[14]. The ratio of the maximum to minimum number of TEACH in Economics, History, and Physics is respectively 21.6, 6.4, and 25.8. Physics students receive the largest mean TEACH (74.6 over a 3-year degree), approximately 2.3 times that of History (32.6 TEACH) and 2.9 times that of Economics

[14] This is even after omitting the five universities that offered the highest number of TEACH. Removing these 'outliers' makes the diagram clearer.

(26.1 TEACH)[15]. Although measures of TEACH are useful in ranking resources provided by different universities, how these data are presented to prospective students 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 lectures (100 students) and 6 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 tradeoffs. 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,

[15] The median number of TEACH per degree for Physics, History and Economics are 46.2, 31.5 and 20.4 respectively.

students might prefer an extra contact hour more than the required reduction in class size. In earlier work 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 complementary between teaching intensity and independent study (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? This shows 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 TEACH fails to capture what students care about or that students can’t benchmark the teaching they receive (Brown et al. (2015)).

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

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

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[16] and location[17], and ϵ_{ijk} is an error term. A log-linear model was chosen because we would expect diminishing marginal effects of degree- and university-specific variables on TEACH.

We have also included a dummy variable for the Russell Group[18], a collection of prestigious public research universities.

In the second set of models we also 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:

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

where i represents degree, j represents university and k represents the department (it captures all courses offered by the Economics department, for example). X is the same

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

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

[18] 11 of the universities in our dataset are members of the Russell Group. We tried controlling for membership of other groups, such as the University Alliance, but there were problems of multi-collinearity.

vector as before, with the addition of TEACH and other characteristics that might affect NSS scores (e.g. use of hourly- paid staff, the proportion of small class teaching and the proportion of part- time students), *Y* is the university-specific variables included in *Z*, and *W* is the department-level variables (e.g. proportion of female students, and proportion of white students) collected from the heidi database[19]. As well as the controls used in Equation 2, this model includes degree-level variables such as the proportion of students who obtained a first class honours, and the proportion of classes with fewer than nine students. All of this information was included separately to allow for independent effects on NSS scores but that 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 (Gibbons et al. (2013)), and unreliable as a measure of education quality (Brown et al. (2015)).

---

[19] heidi is a database run by HESA, and is a source of quantitative data on higher education providers.

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 to Economics and History. Thus, for this regression, we split the data in Table 3: Column (a) presents the results for Physics, and Column (b), for Economics and History. The tables present the results of Equation 2, with only the location variables not shown.

Membership of the Russell Group is a significant predictor of TEACH per degree, with the coefficient dramatically different for Physics compared to Economics and History. Table 3 Column (a) 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 Table 3 Column (b), which suggest that Russell Group universities provide 38 per cent fewer TEACH per degree to Economics and History students, compared to non-Russell Group universities. Thus, differences between subjects are largest amongst the Russell Group.

We also find that TEACH are increasing in research strength for all subjects[20]. However, conditional on Russell Group membership, there is a weak positive correlation between endowment (measured in £m) and TEACH per degree for Economics and History, suggesting that wealthy universities can 'afford' better arts, humanities and social science teaching.

The coefficient on fees is close to zero and insignificant for all three subjects, but negative for Economics and History students. How much students must pay in tuition fees makes no difference to how much teaching they receive. Prima facie, these findings are hard to reconcile with the rationale for variable tuition fees put forward in the 2003 and 2011 white papers (DfES (2003) and BIS (2011)). This variation in TEACH 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 determinates of TEACH (using variables which 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.

---

[20] 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 SD of 54.

TABLE 3

TEACH per degree

	(a) Physics	(b) Economics and/or History
Hourly paid staff used for classes	-0.382	0.147
	(0.219)	(0.0941)
Cohort size	-0.00168	-0.00134***
	(0.00116)	(0.000341)
Russell Group	0.599*	-0.384*
	(0.261)	(0.167)
Fee (£'000)	0.0689	-0.361***
	(0.166)	(0.0835)
Endowment (£m)	-0.00943***	0.00294
	(0.00209)	(0.00171)
Research strength	0.00874***	0.00520***
	(0.00120)	(0.00136)
Observations	75	207
R^2	0.739	0.359

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Location is included in the regression but not displayed

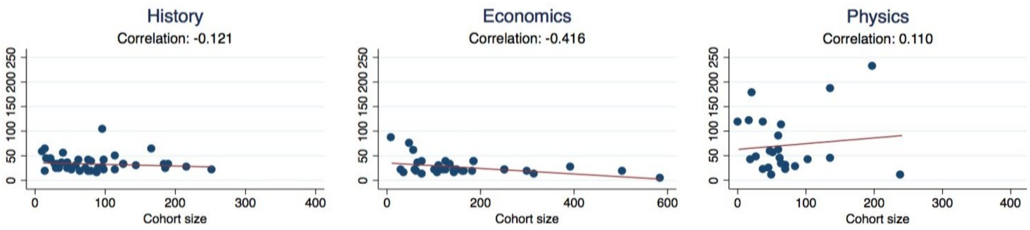

FIGURE 5

Relationship between cohort size and TEACH

2. Does TEACH affect student satisfaction?

This section 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 if teaching 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 survey contains 22 questions, split into five categories, and an overall satisfaction score. The categories are: (i) teaching; (ii) assessment and feedback; (iii) academic support; (iv) organisation; management and resources; and (v) personal development. In Appendix 4 we provide summary statistics for these categories split by subject, and run regressions with the averages of the these as the dependent variables.

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 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 eight and six percent respectively). The co efficient on fees suggests that a £1,000 increase in fees is associated with a reduction in overall satisfaction of 2.4 percent, 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 separately (Appendix 5), especially the relative dissatisfaction of Economics and History students. However, in most of these categories TEACH per degree were 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 VII.

TABLE 4

Overall satisfaction

	Overall satisfaction
TEACH per degree	-0.0000698
	(0.0000527)
Proportion of classes that are 'small'	0.0206
	(0.0381)
Cohort size	0.0000470
	(0.0000515)

History	-0.0615
	(0.0364)
Economics	-0.0791*
	(0.0320)
Hourly paid staff used for classes	0.00464
	(0.0142)
Hourly paid staff used for lectures	-0.0204
	(0.0147)
Hourly paid staff used for practicals	-0.0745***
	(0.0170)
Fee (£'000)	-0.0239
	(0.0180)
Russell Group	-0.0249
	(0.0198)
Endowment (£m)	-0.000468*
	(0.000204)
Research strength	0.000187
	(0.000132)
Prop of students with first class honours	0.000436
	(0.000504)
Proportion of part-time students	0.169***
	(0.0445)
Proportion of female students	0.220*
	(0.0978)
Proportion of disabled students	-0.388
	(0.225)
Proportion of white students	0.118
	(0.0704)
Observations	219
R^2	0.422

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Location is included in the regression but not displayed

These findings raise questions about the use of NSS scores as a proxy for learning gain. Learning gain in education is difficult to measure, 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 student outcomes can be observed by parents, teachers and researchers (Arum and Roksa (2011)). 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 necessary 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 and 2000, and its implications for the unit of resource, staff student ratios, and class size, formed the core of the case for the introduction of tuition fees (Greenaway and Haynes (2003)). Therefore if TEACH are 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

In this section we explain the results of three robustness checks. Firstly, we experimented with adjusting the weighting used for calculating EACHs (see Section VI.). When using both weightings of $w_n = \sqrt{n}$ and $w_n = n^2$, the magnitude of the coefficients naturally altered, but the signs and significance were unchanged. Secondly, w e ran the NSS regressions (see Section VII.2.) using only the proportion of students who answered *'strongly agree'* to each question. The overall picture was unchanged, with only some coefficients increasing in magnitude and a few insignificant coefficients changing signs. Finally, 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 in Economics and History practicals are on average much larger than classes, whilst in Physics they are a similar size.

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 reveals large differences in the 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 receive almost five times as much teaching as students in the bottom decile. We also find large differences between the three subjects: 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 (Britton et al. (2015)), which may result from signaling or human capital (Weiss (1995)). The optimal amount of TEACH per degree will depend on which hypothesis you subscribe to[21]. 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.

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

Some subjects receive a subsidy from the treasury depending on their price band classification by HEFCE. The standard explanation for these differences is that STEM subjects have higher fixed costs (laboratory equipment, etc.), than arts, humanities and social science (HEFCE (2005))[22]. It has always been the case that students studying STEM subjects had more contact hours but in the past, this was offset by larger classes (Robbins (1963)). 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[23]. 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: students studying Physics at a Russell Group university are

[22] The latter are often referred to a library based disciplines. See for example this plea for Economics to be reclassified http://www.res.org.uk/view/art3Apr04Features.html

[23] Physics is a band B subject and in 13/14 received £1500pa extra from HEFCE on top of the £9000 fee paid by students. Economics and History are both band D, and receive zero teaching subsidy (HEFCE (2014)). Thus the unit of resource for Physics (£9000+£1500) is 116.7% of that for Economics or History. We find mean TEACH for physics is 230% of Economics and 290% of History.

advantaged, whilst students studying Economics or History are penalised. However, little of the variation in teaching can be explained by factors such as the fee, research reputation, or endowment: accident and history appear to determine the outcome. We found that NSS results cannot be satisfactorily explained by either the existing data (e.g. endowment or REF score) or our new administrative data (e.g. 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 (Brown et al. (2015)).

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 in the market for teaching price signals are weak, and that this raises important issues for the design of policy. This does not imply that market forces in higher education are absent. 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 his recent announcement Johnson (2015) stated: *‘students must be provided with clearer information about how many hours students will spend in lectures, seminars and tutorials,* *and who will deliver the teaching'*. Our framework allows a student choosing between courses to focus on three questions: (i) the units of teaching on offer (i.e. the TEACH); (ii) how these resources are allocated across different learning activities (i.e. lectures, classes, practicals); and (iii) 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 exists which 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 departmental 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 they collected. Given all the advances in data recording, storage and analysis that have taken place since 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 TEF 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 REF, 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 these in increasingly high quality. In higher education the critical allocation problem is between teaching and research. Our metric is unique in its focus on this margin and we believe if adopted it will help re-balance 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 HE and the outcomes they generate. Undertaking the necessary research can only happen if universities are transparent about the teaching they deliver.

References

Allgood, S., Walstad, W. B., and Siegfried, J. J. (2015). Research on teaching economics to undergraduates. *Journal of Economic Literature*, 53(2):285–325.

Arias, J. and Walker, D. M. (2004). Additional evidence on the relationship between class size and student performance*. The Journal of Economic Education,* 35(4):311–329.

Arum, R. and Roksa, J. (2011). Academically adrift: Limited learning on college campuses. *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,* 20(1):43–59.

Bandiera, O., Larcinese, V., and Rasul, I. (2010). Heterogeneous class size effects: New evidence from a panel of university students*. *The Economic Journal,* 120(549):1365–1398.

Barr, N. (2004). Higher education funding. *Oxford review of economic policy,* 20(2):264–283.

Becker, W. E. and Powers, J. R. (2001). Student performance, attrition, and class size given missing student data. *Economics of Education Review,* 20(4):377–388.

BIS (2011). Higher Education: Students at the Heart of the System, volume 8122. *The Stationery Office*.

BIS (2016). Success as a Knowledge Economy: Teaching Excellence, Social Mobility and Student Choice, volume cm9258. *The Stationery Office*.

Britton, J., Shephard, N., and Vignoles, A. (2015). Comparing sample survey measures of english earnings of graduates with administrative data during the great recession. Technical report, *IFS Working Papers.*

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 models. *Journal of behavioral decision making*, 28(1):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*, 36(6):693–712.

De Fraja, G. and Valbonesi, P. (2012). The design of the university system. *Journal of Public Economics*, 96(3):317–330.

DfE (2016). Teaching Excellence Framework: year two and beyond. *The Stationery Office*.

DfES (2003). The future of higher education.

GB (2000). Freedom of information act 2000. London: *The Stationery Office*.

Gibbons, S., Neumayer, E., and Perkins, R. (2013). Student satisfaction, league tables and university applications.

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

Gibbs, G. (2012). Implications of dimensions of quality in a market environment. *York: Higher.*

Greenaway, D. and Haynes, M. (2003). Funding higher education in the uk: The role of fees and loans*. *The Economic Journal*, 113(485): F150–F166.

Hale, E. and Tattersall, A. (1964). University teaching methods: A report of the committee on university teaching methods.

HEFCE (2005). Review of the teaching funding method.

HEFCE (2014). Guide to funding and student number controls 2013-14 and 2014-15: How hefce allocates its funds and controls student numbers.

HEPI (2016). Hepi student academic experience survey.

Hillman, N. (2015). Data-led policy: What should the 2015 hepi / hea academic experience survey mean for policy?

Hoxby, C. M. (2014). The economics of online postsecondary education: Moocs, nonselective education, and highly selective education. Technical report, *National Bureau of Economic Research.*

Huxley, G. and Peacey, M. W. (2014). Modelling learning parameters. Technical report, Department of Economics, University of Bristol, UK.

Huybers, T., Louviere, J., and Islam, T. (2015). What determines student satisfaction with university subjects? a choice-based approach. *Journal of choice modelling*, 17:52–65.

Johnson, J. (2015). Teaching at the heart of the system. Speech, page 1 June.

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,* 27(2):221–233.

Krueger, A. B. (2003). Economic considerations and class size*. *The Economic Journal*, 113(485): F34–F63.

Lenton, P. (2015). Determining student satisfaction: An economic analysis of the national student survey. *Economics of Education Review,* 47:118–127.

Martins, P. S. and Walker, I. (2006). Student achievement and university classes: Effects of attendance, size, peers, and teachers.

Monks, J. and Schmidt, R. M. (2011). The impact of class size on outcomes in higher education. *The BE Journal of Economic Analysis & Policy*, 11(1).

Richardson, J. T., Slater, J. B., and Wilson, J. (2007). The national student survey: development, findings and implications. *Studies in Higher Education*, 32(5):557–580.

Robbins, L. R. B. (1963). Higher Education: report of the committee appointed by the Prime Minister under the chairmanship of Lord Robbins,1961-63. Number 2. *HM Stationery Office*.

Salop, S. and Stiglitz, J. (1977). Bargains and ripoffs: A model of monopolistically competitive price dispersion. *The Review of Economic Studies*, pages 493–510.

Sapelli, C. and Illanes, G. (2016). Class size and teacher effects in higher education. *Economics of Education Review*, 52:19–28.

Schanzenbach, D. (2010). The economics of class size. *Economics of education*, pages 183–190.

Weiss, A. (1995). Human capital vs. signalling explanations of wages. *The Journal of Economic Perspectives*, 9(4):133–154.