# Heterogeneous Class Size Effects:
# New Evidence from a Panel of University Students[*]

Oriana Bandiera[†]       Valentino Larcinese[‡]       Imran Rasul[§]

October 2009

## Abstract

Over the last decade, many countries have experienced dramatic increases in university enrolment, which, when not matched by compensating increases in other inputs, have resulted in larger class sizes. Using administrative records from a leading UK university, we present evidence on the effects of class size on students' test scores. We observe the same student and faculty members being exposed to a wide range of class sizes from less than 10 to over 200. We therefore estimate non-linear class size effects controlling for unobserved heterogeneity of both individual students and faculty. We find that – (i) at the average class size, the effect size is $-.108$; (ii) the effect size is however negative and significant only for the smallest and largest ranges of class sizes and zero over a wide range of intermediate class sizes; (iii) students at the top of the test score distribution are more affected by changes in class size, especially when class sizes are very large. We present evidence to rule out class size effects being due solely to the non-random assignment of faculty to class size, sorting by students onto courses on the basis of class size, omitted inputs, the difficulty of courses, or grading policies. The evidence also shows the class size effects are not mitigated for students with greater knowledge of the UK university system, this university in particular, or with greater family wealth.

Keywords: class size, heterogeneity, university education.

JEL Classification: A20, D23, I23.

[†]Department of Economics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom; Tel: +44-207-955-7519; Fax: +44-207-955-6951; E-mail: o.bandiera@lse.ac.uk.

[‡]Department of Government, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom; Tel: +44-207-955-6692; Fax: +44-207-955-6352; E-mail: v.larcinese@lse.ac.uk.

[§]Department of Economics, University College London, Drayton House, 30 Gordon Street, London WC1E 6BT, United Kingdom. Telephone: +44 (0)20 7679 5853; Fax: +44 (0)20 7916 2775; E-mail: i.rasul@ucl.ac.uk.

# 1　Introduction

The organisation of university education remains in the spotlight both in academia and policy circles. Recent research has stressed the importance of higher education in providing positive externalities within firms [Moretti 2004], within local labour markets [Glaeser *et al* 1992], and in fostering economy wide growth as a whole [Aghion *et al* 2007]. Concurrently, most OECD countries have adopted supply side policies that have led to dramatic increases in university enrolment during the last decade.[1]

To illustrate, undergraduate enrolment in the UK has risen by over 30% since 1995 [Figure 1A]. Over the same period, postgraduate enrolment has increased even more dramatically while the number of faculty has grown at a much slower pace [Figure 1B]. To the extent that universities cannot instantaneously adjust all relevant inputs in response to such increases in enrolment – such as the number of faculty or teaching rooms – students inevitably face larger class sizes.

The effect of increasing class size in tertiary education is not well understood. While the established literature on class size effects in primary and secondary schools provides useful guidance, further investigation is needed in the context of university education as the range of class sizes is typically larger than at other tiers of the education system, the underlying sources of variation in class size used for identification differ, and the mechanisms that drive class size effects are likely to differ as well.

Although tertiary education may involve more self-learning than primary or secondary education, class size is solidly at the top of the policy agenda and concerns of both faculty and students. This is particularly evident in the UK where concerns on the increasing student-to-staff ratios in higher education institutions have recently been expressed in a report of the Department for innovation, universities and skills[2] and by the most important unions of university teachers.[3] The student-to-staff ratio is a also a commonly used indicator of quality both in national and international comparisons. For instance, the student-to-staff ratio accounts for 20% of the scores that determine the global ranking of higher education institutions by the Times Higher Education Supplement.[4] It is used for the same purposes by three leading UK broadsheets (The Guardian, The Independent and The Times) for their yearly ranking of UK universities and by the US News and World Report for their ranking of American colleges.[5]

Not surprisingly, the issue of class size in higher education has also attracted the attention of policy makers. The landmark 1997 report of the National Committee of Inquiry into Higher Education (informally

---

[1]Within the OECD, the UK is actually at the low end of enrolment growth. For example, between 1998 and 2005 the US experienced a 30% increase in student enrolment. The corresponding figure for the UK was 18%. Finland, Ireland, the Netherlands, and Sweden have all experienced greater proportionate increases in enrolment.

[2]The report "The future of higher education" published in December 2008 is authored by the head of the Higher Education Academy, Paul Ramsden.

[3]In the paper "Student retention - problems and solutions" (June 2001), the Association of University Teachers includes "the doubling of student staff ratios over the past two decades" among the possible causes of increasing drop out rates. In the paper "Further, Higher, Better" of September 2006, the University and College Union provides extensive evidence of growth in the student-staff ratio and states to be "extremely concerned" about it.

[4]This choice is justified by the fact that the "student faculty ratio is, at present, the only globally comparable and available indicator that has been identified to address the stated objective of evaluating teaching quality." (http://www.topuniversities.com/worlduniversityrankings/methodology/studentfaculty, accessed on July 24, 2009).

[5]The Times uses 9 indicators, the Independent 8, the Guardian 7. In the case of the Guardian, the average student-to-faculty ratio is the indicator that receives the highest weight (20%) in the aggregate score.The U.S. News and World report ranking is based on 16 indicators. Faculty resources is worth 20% of the total score and average student-to-faculty accounts for 40% of the faculty resoures indicator

known as Dearing's Review), which was commissioned by the government and led to the introduction of tuition fees in the UK, explicitly states that "students perform worse in large classes".

Despite its policy relevance, evidence on the causal effect of class size on students' performance is limited. This paper analyses the administrative records from a leading UK university to identify the causal impact of class size on the academic achievement of postgraduate students – the UK equivalent of US college seniors.[6] There are several mechanisms through which class size can affect the behaviour of students and faculty. These behavioural changes can occur inside and outside the lecture theatre. For example, students may be less attentive in larger classes, or may compensate for larger classes by exerting more effort either in the library or with their peers. Faculty may be better able to identify the ability and interests of the median student in smaller classes, or be more able to answer students' questions directly. Outside of the lecture theatre, faculty might devote more time preparing the delivery of lectures and organisation of materials for larger classes, or there may be congestion effects if faculty have less time to devote per student during office hours.[7]

We measure academic achievement through the student's end of year final exam performance. Test scores are a good measure of student's performance and learning both because they are not curved, and hence measure the students' absolute performance, and because, unlike in many North American universities [Hoffman and Oreopoulos 2006], faculty have neither the incentive nor the possibility to strategically manipulate test scores to boost student numbers or to raise their own student evaluations. Class size is measured as the number of students enrolled to take the end of year exam. We deliberately focus on identifying the causal effect of student enrolment, as opposed to attendance, because enrollment is a policy parameter that universities can measure and manipulate relatively easily. In contrast, it is orders of magnitude more costly for universities to measure and regulate the physical attendance of each student in each of their classes.

Our administrative data has three key features that we exploit to identify the effect of class size on academic achievement. First, each student takes on average four courses within the same subject area and there is variation in the class sizes across courses they are exposed to. This allows us to present within student estimates of class size controlling for individual time invariant characteristics that equally affect performance across courses such as the individual's underlying ability, employment options upon graduation, and past educational investments.

Second, there is considerable within student variation in class size. The median student has a difference between her largest and smallest class sizes of 56, and more than 20% of students have a difference of at least 100. We therefore estimate whether class size effects are non-linear over a wide range of class sizes,

---

[6]The UK higher education system awards three types of degree. Undergraduate degrees are awarded upon completion of a three year course, postgraduate (or masters') degrees after one further year of study, and doctoral degrees upon completion of further courses and research.

[7]There is mixed evidence from primary schools on whether student behaviors directly related to learning – such as attentiveness – are affected by class size [Finn *et al* 2001]. There is more consistent evidence that in smaller classes, disruptive behavior decreases and collaborative behavior increases [Johnston 1990, Blatchford *et al* 2005], which is in line with the theoretical predictions of Lazear [2001]. In university settings, there is evidence that students attitudes towards learning tends to be negatively affected by larger classes [Bolander 1973, Feldman 1984, McConnell and Sosin 1984]. On teacher behavior, there is evidence from primary school settings that teachers know students better in smaller classes [Johnston 1990, Boyd-Zaharias and Pate-Bain 2000, Blatchford *et al* 2005] but evidence on faculty behavior across class sizes in university setting remains scarce.

from less than 10 to over 200. This is important given some potential mechanisms for class size effects are only relevant, on the margin, in the smallest or largest classes.

Third, we observe the same faculty member teaching different class sizes. We are thus able to control for faculty fixed effects, which capture the faculty member's teaching style or motivational skills. We use this aspect of the data to infer how faculty may be altering their behaviour as they teach courses of very different size.

Our main results are as follows. First, the baseline within student and teacher estimates imply a negative and significant effect size of $-.108$. This implies that if a student were to be reassigned from a class of average size to a class which was one standard deviation larger, the test score of the same student would fall by .108 of the within student standard deviation in test scores.

Second, the effect of class size on test scores is highly non-linear across the range of class sizes we observe. More precisely, we find a large negative effect going from small (1-19) to medium (20-33) class sizes, a close to zero effect for further increases over a wide range of intermediate class sizes – corresponding to the third and fourth quintiles of class size in our data (34-103), and an additional negative effect in the top quintile of class sizes (104-211). Our findings are consistent with the only other study that identifies the causal impact of class size on university test scores. Machado and Vera Hernandez [2008] exploit the quasi-random variation in class size induced by the assignment of students to classes on the basis of their surname in a leading Spanish university, and find no effect of class size on grades over a range of class sizes from 50 to 70.[8]

That the effect of class size is non-linear suggests that there are at least two underlying mechanisms at play, one that explains the effect at small class sizes, and one for the largest class sizes. The finding also helps rule out that class size is capturing other omitted inputs. *A priori*, most omitted inputs, such as the audibility of the lecturer or congestion effects outside of the lecture theatre, should indeed be linearly related to class size. To the extent that a robust non-linear relationship between class size and test scores exists, omitted inputs are unlikely to explain the full pattern of class size effects.

Third, using quantile regression methods we find that in the smallest classes, namely less than 33, larger class sizes uniformly reduce conditional test scores of all students. In the intermediate range of class sizes on which the effect size is zero, the quantile regression estimates also confirm that there are no class size effects on the distribution of test scores as a whole. However, the quantile regression estimates show that increases in class size when classes are in the top quintile of the class size distribution to begin with, significantly reduce tests scores and have a greater detrimental effect on test scores at the top end of the distribution. This suggests there exists an important complementarity between student ability and class size. The highest ability students would benefit the most, in terms of academic achievement, from any reduction in class sizes, when class sizes are very large to begin with.[9]

---

[8]Our estimates are at the low end of previously documented class size effects in primary and secondary education settings. Krueger [1999] uses data from Project STAR that randomly assigned teachers and students to different class sizes. Using student level estimates with school fixed effects he reports an effect size of $-.20$ for kindergarten and $-.28$ for first grade. Angrist and Lavy [1999] use Maimonides rule in Israeli schools as the basis of an IV for class size among 3rd to 5th graders, where class sizes are capped at and the median class size is 31. They find an effect size of $-.29$ in their class level analysis. Hoxby [2000] uses variations in class size driven by idiosyncratic variation in population size in class sizes of 10 to 30. She finds a zero effect size in her class level analysis. Finally, Duflo *et al* [2007] use a randomized field experiment to reduce class sizes in Kenyan primary schools from 80 to 46 – they find little effect on test scores but large reductions in teacher effort.

[9]Arulampalam et al [2007] find that also the effect of attendance is stronger for higher ability students.

Fourth, we use information on teachers' assignments to classes and on students' characteristics to shed light on the underlying mechanisms for the class size effect. We find no evidence that departments purposefully assign faculty of differing quality to different class sizes. We also find no evidence that faculty members alter their behaviour when exposed to different class sizes, consistent with the preparation and delivery of lectures being independent of the number of students taught.

Finally, we find the class size effect does not vary with proxies for students' wealth. This casts doubt on the relevance of infrastructure congestion in explaining the documented class size effect, namely if larger classes resulted in lower grades because students had more limited access to library books or computer laboratories, the effect should have been smaller for students who can purchase these inputs privately. Moreover, we also find the class size effect does not vary with student's familiarity with this particular university or with the UK system in general. This casts doubts on the relevance of mechanisms that work through the information available to students, such as their awareness of other local resources, such as other libraries in the area, or their knowledge of the characteristics of faculty, courses, or departments.

Against a backdrop of rapidly increasing enrolment rates in the UK – as in much of the OECD – our analysis has important policy implications for university education. While there is robust evidence of a negative class size effect on the academic achievement of students, we also document there exists a wide range of class sizes over which, on the margin, targeting resources to reduce class sizes will have little impact on test scores. This suggests that in this range it would be more efficient to spend resources on other inputs. However, eliminating the largest classes– namely those over 100 – not only will raise test scores, it will do so to a greater extent for the most able students.

The paper is organised as follows. Section 2 describes the empirical context, data sources, and presents descriptive evidence on class sizes and test scores. Section 3 describes the empirical method. Section 4 presents our baseline class size effect estimates and addresses econometric concerns. Section 5 presents evidence on whether these effects are heterogeneous across class sizes and quantile regression estimates of whether marginal changes in class size have heterogeneous effects across students. Section 6 presents evidence on some mechanisms that might be driving the class size effect. Section 7 discusses the external validity of our findings and their policy implications. Further results and robustness checks are in the Appendix.

# 2    Context and Data

## 2.1    *Institutional Setting*

Our analysis is based on administrative data on individual students from a leading UK university, for the academic years 1999/00 to 2003/04. The UK higher education system comprises three tiers – a three-year undergraduate degree, a one or two-year M.Sc. degree and Ph.D. degrees of variable duration. We focus on full time students enrolled on one-year M.Sc. degree programs. These students will therefore have already completed a three year undergraduate degree program at some university.[10]

---

[10]Students are not restricted to only apply to M.Sc. degree programs in the same field as that in which they majored in their undergraduate degree. In our sample, 17% of students studied for their undergraduate degree at the same institution, and around 17% have previously studied in the UK.

Over the academic year each student must obtain a total of four credits, one of which is assigned upon completion of a long essay, and the remaining three upon the completion of final examinations related to taught courses. As each course is worth either one or half a credit, the average student takes 3.7 taught courses. Taught courses are assessed through an end of year sit down examination and all final examinations take place over the same two week period in June.[11]

The university has 23 academic departments, offering 125 degree programs in total. Students enrol onto a specific degree program and once enrolled, they cannot move to other programs or departments. Each degree program has its own associated list of core courses – that are either compulsory or offering a very constrained choice – as well as a list of elective courses. The latter might also include courses belonging to other programs or departments. For instance, a student enrolled in the M.Sc. degree in economics can choose between a basic and an advanced version of the core courses in micro, macro, and econometrics plus an elective course from a list of economics fields and a shorter list from other departments such as finance. The logistics and teaching for each course are the responsibility of a specific department. There is typically only one section for each course, so students do not have the possibility to sort into smaller sections of the same course. Only a very small percentage of students enrolled on M.Sc. degree programmes continue to Ph.D. level studies in the same institution.

## 2.2 Data Sources

The university's administrative records contain information on each student's academic performance on each course, as well as some individual background characteristics. Our sample covers 10,873 students enrolled full-time on one year M.Sc. programs over academic years 1999/00-2003/04, on 626 different courses. There are a total of 40,851 student-course level observations and the primary unit of analysis is student $i$ in course $c$ in academic year $t$. The administrative records identify all students enrolled on the same course in the same academic year. Hence we are able to construct measures related to the composition of the class on the basis of demographic characteristics as well as study related characteristics such as the fragmentation of students across different enrolment departments, and whether the course is a core or elective course for any given student. Finally, we coded university handbooks as these provide information on the assignment of faculty to each course in each academic year. We identify 794 teaching faculty and their rank – assistant, associate or full professor.

## 2.3 Key Variables

### 2.3.1 Test Scores

Our main outcome variable is the final exam performance of student $i$ on course $c$, $y_{ic}$, scaled from 0 to 100. This translates into the final classmarks as follows – grade A corresponds to test scores of 70 and above, grade B corresponds to 60-69, grade C to 50-59, and a fail to 49 or lower.[12]

---

[11]Some courses are also partially assessed through coursework – for such courses, typically no more than 25% of the final mark stems from the coursework component. Throughout the empirical analysis we focus on examined courses and control for a set of course characteristics.

[12]Despite degree programs in this university being highly competitive to enter and students paying amongst the highest fees of any UK university, we still observe students failing courses – 3% of observations at the student-course-year level

In this institutional setting final exam scores are a good measure of student's performance and learning for a number of reasons. First and foremost, test scores are not curved so they reflect each individual's absolute performance on the course. Grading guidelines issued by the departments indeed illustrate that exam grading takes place on an absolute scale.[13] Figure 2 provides further evidence on the fact that grades are not curved. Each vertical bar represents a particular course in a given academic year, and the shaded regions show the proportion of students that obtain each classmark (A, B, C, or fail). For ease of exposition, the course-years are sorted into ascending order of B-grades, so that courses on the right hand side of the figure are those in which all students obtain a test score of between 60 and 69 on the final exam. We note that, in line with marks not being curved, there exist some courses on which *all* students obtain the same classmark. On some courses this is because all students obtain a B-grade, and on other courses all students achieve an A-grade. In 23% of course-years, not a single student obtains an A-grade. In addition, the average classmark varies widely across course-years and there is no upper or lower bound in place on the average grade of students in any given course-year. We later present evidence that the identity of faculty teaching the course matters for the exam performance of students, again contrary to grading on a curve.[14]

Second, unlike in many North American universities, faculty do not have incentives to manipulate the exam scores or course content to boost student numbers or to raise their own student evaluations [Hoffman and Oreopoulos 2006]. In line with this, we note that smaller courses are not systematically more likely to be discontinued, all else equal.[15] Moreover, manipulating test scores is difficult as exam scripts are double blind marked by two members of faculty (and sent to a third marker, external to the university, in case of disagreement over the final score), of which typically only one teaches the course.

### 2.3.2 *Class Sizes*

We define class size as the number of students formally enrolled to take the end of year exam. This can diverge from number of students *physically* present in the lecture theatre because not all students who are enrolled to take the exam necessarily attend the classes. While we do not have direct evidence on attendance, the fact that students in this university pay among the highest annual fees in the UK, suggests that the majority of students are likely to attend, in line with evidence from elite universities in the US and Spain [Romer 1993, Machado and Vera Hernandez 2008].[16]

---

correspond to fails. The incidence of dropping out – namely students enrolling onto programs and not sitting exams – is very low. Finally, all exams take place at the end of the academic year so students are not selected out of the sample, by having failed mid terms for example.

[13]For example, the guidelines for one department state that an A-grade will be given on exams to students that display "a good depth of material, original ideas or structure of argument, extensive referencing and good appreciation of literature", and that a C-grade will be given to students that display "a heavy reliance on lecture material, little detail or originality".

[14]An alternative check on whether test scores are curved is to test whether the mean score on a course-year differs from the mean score across all courses offered by the department in the same academic year. For 29% of course-years we reject the hypothesis that the mean test score is equal to the mean score at the department-year level. Similarly for 22% of course-years we reject the hypothesis that the standard deviation of test scores is equal to the standard deviation of scores at the department-year level.

[15]More precisely, we find that the existence of course $c$ in academic year $t$ does not depend on enrolment on the course in year $t-1$, controlling for some basic course characteristics and the faculty that taught the course in $t-1$.

[16]In addition to lectures, students are required to attend smaller tutorials where problem sets are worked through. In many cases Ph.D. students run tutorials. The focus of this paper is not to test whether tutorial sizes or attendance affect

Moreover, we deliberately focus on identifying the causal effect of student enrolment, as opposed to student attendance, on exam performance because enrollment is a policy parameter that universities can measure and manipulate relatively easily. In contrast, it is more costly for universities to measure and regulate the physical attendance of each student in each of their lectures. In addition, the number of students enrolled on the course captures the competition each student faces for resources both inside lectures – such as the ability to hear the teacher or ask her a question – as well as competition for resources outside of lectures – such as library books and access to faculty during office hours.

It is instructive to understand how our parameter of interest relates to the effect of the number of students that physically attend classes on test scores. Consider the simple linear regression model in which we regress the test score of student $i$ in class $c$ on the number of students physically present in the class, $Z_c$, $y_{ic} = \beta_p Z_c + u_{ic}$, where the error term is assumed classical. This yields the estimated effect of physical presence on test scores, $\widehat{\beta}_p$. In our analysis we regress test scores on class enrolment, $N_c$, $y_{ic} = \beta_e N_c + e_{ic}$, where the error term is again assumed classical. Assuming a relationship between class enrolment and class attendance of the form $N_c = Z_c + w_c$ enables us to identify the conditions under which the class size effect parameter we estimate based on enrolment into the course, $\widehat{\beta}_e$, is smaller in absolute value than the effect of physical attendance on the course on test scores, $\widehat{\beta}_p$.

Under the reasonable assumption that the gap between enrollment and the number of students that physically attend is positively correlated with attendance, a sufficient condition for $\left|\widehat{\beta}_e\right| \leq \left|\widehat{\beta}_p\right|$ when $\widehat{\beta}_p < 0$ is that the wedge between attendance and enrollment ($w_c$) is either uncorrelated or positively correlated with other unobserved determinants of test scores. Thus, if students are less likely to attend courses for which test scores are exogenously higher – for instance due to the course material being easy – our estimate $\widehat{\beta}_e$ provides a lower bound on the effect of physical attendance on the course on test scores, $\widehat{\beta}_p$. The opposite can occur only if $cov(w_c, y_{ic})$ is negative and sufficiently large.[17] Of course more complicated models can be written to describe how physical attendance, enrolment, and test scores interrelate. The point to emphasise is that there are reasonable conditions under which our class size effects based on enrolment provide a lower bound for the effect of physical attendance on test scores. More importantly, enrolment is a policy relevant parameter whereas physical attendance of every student to every class is more costly for universities to control.

# 3 Descriptives and Empirical Method

## 3.1 *Descriptives*

Table 1 presents descriptive evidence on our two key variables – test scores and class sizes. The average test score is 62.0, the standard deviation in test scores between students is 5.00, and most relevant for our

---

exam performance because in this institution, many tutorial sizes are capped at 15 by Maimonides rule, so there is little variation to exploit. Martins and Walker [2006] find no effect of tutorial size on test scores for economics undergraduates at the University of Warwick.

[17]To derive these conditions consider that $\widehat{\beta}_e = \frac{cov(Z_c, y_{ic}) + cov(w_c, y_{ic})}{var(Z_c + w_c)}$ and $\widehat{\beta}_p = \frac{cov(Z_c, y_{ic})}{var(Z_c)}$. Moreover $cov(Z_c, w_c) > 0 \Rightarrow var(Z_c + w_c) > var(Z_c)$. Thus if $cov(Z_c, y_{ic}) \leq 0, cov(w_c, y_{ic}) \geq 0 \Rightarrow \left|\widehat{\beta}_e\right| \leq \left|\widehat{\beta}_p\right|$. If $cov(Z_c, y_{ic}) > 0$, so that the effect of attendance on test scores $\widehat{\beta}_p$ is positive, the conditions are reversed.

analysis, the within student standard deviation in test scores is almost as large, 4.37. These test score statistics are very similar when we consider core and electives separately. We define the within student test score gap as being the difference between student $i$'s highest and lowest test score across all her examined courses. Figure 3A then shows a histogram of the within student test score gap. The median individual has a test score gap of 9, and around 20% of students have a test score gap of at least 15. In short, there is considerable variation in test scores within the same student to explain.

The remainder of Table 1 shows that the average class size is 56.2, and the within student standard deviation in class size is 32.3 – comparable in magnitude to the standard deviation in class sizes between students, 33.2. Core courses tend to have larger class sizes than electives, although within each course type, the same student is exposed to enormous variation in class sizes, despite courses being in the same subject area. We define the within student class size gap as the difference between student $i$'s largest and smallest class size across all her courses. Figure 3B then shows a histogram of the within student class size gap. The median individual has a class size gap of 56, and more than 20% of students have a class size gap of at least 100, and so are exposed to class sizes in all five quintiles of the class size distribution.[18]

The underlying basis for our analysis is that universities cannot instantaneously adjust on all margins in response to increases in student enrolment, as documented in Figure 1. If universities and departments could costlessly adjust all inputs and they aimed to maximise test scores, then an envelope theorem argument would imply class sizes would be set optimally and there would be no effect on the margin of class size adjustments on test scores [Lazear 2001]. In the Appendix we provide evidence in support of the fact that departments cannot fully adjust inputs in response to changes in enrolment over time. As a consequence, increases in enrolment translate into significantly larger class sizes, which in turn suggests plausibly non-zero class size effects can exist in this setting.

## 3.2   Empirical Method

We provide within student estimates of the effect of class size on individual test scores using the following panel data specification,[19]

$$y_{ic} = \alpha_i + \gamma N_c + \delta X_c + \lambda H_c + \sum_j \mu_j f_{jc} + u_{ic}, \tag{1}$$

where $y_{ic}$ is the test score of student $i$ on course $c$, $\alpha_i$ is a fixed effect for student $i$ that captures the individual's underlying ability, motivation, employment options upon graduation, and past educational investments. Exploiting within student variation allows us to control for a number of sources of potential bias. For example, the most able students may sort out of the largest class sizes, which if student fixed effects were not controlled for, would cause $\widehat{\gamma}$ to be downwards biased. Since the effect of class size is identified by comparing the performance of the same student in different courses, it is important to stress

---

[18]Although departments could informally cap class sizes, there are no stated official guidelines on whether and how this is possible. In addition, we do not observe any spikes in the class size distribution at focal values.

[19]As we have no information to be able to condition on past exam performance, we are not estimating a value added model. Such models have been criticized on various grounds [Todd and Wolpin 2003, Rothstein 2007]. Moreover, the type of specification we estimate for the effect of class size on the level of test scores has been argued to better reflect the total effects of class size [Krueger 1999].

that courses belong to the same degree programme and hence cover similar topics, require students to develop similar skills, and use similar methods of assessment.

$N_c$ measures the class size as defined by the number of students enrolled on course $c$. The course controls in $X_c$ capture other determinants of test scores, including the number of faculty that teach on the course, the share of them that are full professors – that may reflect the difficulty of the course, the number of credits obtained for completing the course, whether the class is a core or elective course, and the share of the overall course mark that is attributed to the final exam.

The course controls $H_c$ relate to composition of the peer group in the course. In particular we control for the share of women on the course, the mean and standard deviation of students' ages, the ethnic fragmentation of students, the departmental fragmentation of students, the share of students who completed their undergraduate studies at the same institution, and the share of British students. Controlling for the composition of students on the course, $H_c$, addresses concerns that students on larger courses are likely to have a more heterogeneous group of peers, and a more diverse group of peers may have positive or negative effects on individual test scores.[20]

University handbooks provide information on the assignment of faculty to each course in each academic year. We use this information to control for a complete series of faculty dummies, $\sum_j \mu_j f_{jc}$, such that $f_{jc}$ is one if faculty member $j$ teachers on course $c$, and zero otherwise. These capture factors that cause the academic performance of all students of faculty member $j$ to be affected in the same way, such as the faculty member's teaching style or motivational skills.

Finally, the error term, $u_{ic}$, is clustered by course-academic year to capture common unobservable shocks to students' end of year exam performance such as the difficulty of the final exam script.[21]

In line with the existing literature, the parameter of interest is the effect size, $\widehat{\gamma}\left(\frac{sd(N_c)}{sd(y_c)}\right)$, where both standard deviations are calculated within student because that is the primary source of variation exploited to identify $\gamma$ in (1). Intuitively, the effect size measures the share of the within student standard deviation in test scores that is explained by a one standard deviation increase from the mean class size.[22]

It is important to be precise about the mechanisms that should be captured by our reduced form class size effect estimate, $\widehat{\gamma}$. These relate to any behavioural changes – of students or faculty – that stem from being exposed to different class sizes. The changes in behaviour that we wish to capture can occur either inside or outside the lecture theatre itself. For example, students may be less attentive to faculty lecture delivery in larger classes, or may compensate for larger classes by exerting more effort outside of lecture times, either in the library or with their peers [Bolander 1973, Feldman 1984, McConnell and Sosin 1984]. Inside the lecture theatre, faculty may be better able to identify the ability and interests of the median

---

[20]Students can belong to one of the following ethnicities – white, black, Asian, Chinese, and other. The ethnic fragmentation index is the probability that two randomly chosen students are of different ethnicity. The departmental fragmentation index is analogously defined. We experimented with a number of alternative controls in $X_{ct}$ and $H_{ct}$ – the reported results are robust to small changes in these sets of variables. Estimates of peer effects in university settings have been previously identified using alternative experimental or quasi-experimental empirical strategies [Sacerdote 2001, Zimmerman 2003, Arcidiacono and Nicholson 2005].

[21]We also experimented with alternative forms of clustering such as at the course level, and with weighting the observations by the number of credits the course is worth. The main results are robust to these variations in specification.

[22]To recover the corresponding effect size that is normalized by overall standard deviations, we note that the overall standard deviation in test scores (class sizes) is 6.67 (46.3). As both of these increase in approximately the same proportion over the within student standard deviations, the implied effect sizes will be similar whether the overall or within student standard deviations are used.

student in smaller classes, or be more able to answer students' questions directly. Outside of the lecture theatre, faculty behaviour may be affected if they spend more time preparing the delivery of lectures and organisation of materials for larger classes, or there may be congestion effects if faculty have less time to devote per student during office hours.[23]

# 4 Empirical Results

## 4.1 *Baseline Estimates*

Table 2 presents our baseline estimates of the effect of class size on test scores. Column 1 shows that unconditionally, larger class sizes are associated with significantly lower test scores. The effect size is $-0.074$ and significantly less than zero. Hence, evaluated at the mean, a one standard deviation increase in class size reduces individual test scores by 0.074 standard deviations of the overall distribution of test scores.

Column 2 shows this effect to be robust to conditioning on course-academic year factors, $X_c$. Moreover we note that the magnitude of the class size effect is large relative to other observables controlled for such as the total number of faculty that teach the course, the share of them that are full professors, and the proportion of the overall exam mark that is attributed to the final exam mark.

Column 3 then controls for the student fixed effects, $\alpha_i$. The effect size is slightly larger in absolute value than the earlier specifications, at $-0.082$. This implies that if a student were to be reassigned from a class of average size to a class which was one standard deviation larger, the test score of the same student would fall by 0.082 of the within student standard deviation.[24]

Two points are of note. First, the student fixed effects account for around 56% of the overall variation in test scores. Hence fixed student characteristics such as their underlying ability or motivation to succeed are the single most important determinant of academic achievement. Second, given the similarity of the implied effect size with and without controlling for student fixed effects, the data suggests sorting of students by ability into courses is not an important determinant of the effect of class size on test scores. While we have no doubt that there is a great deal of sorting of students into universities and into degree programs within the same university – just as there is sorting of children into primary and secondary schools – the data from this setting suggests there is no strong element of sorting by students into courses on the basis of class size.[25]

---

[23]How much of any class size effect operates through changes in the likelihood to attend is unclear. As described earlier, our class size effects are likely to underestimate any effect of attendance on enrolment. Studies on the determinants of attendance, and specifically how class size relates to attendance are rare. Romer [1993] presents evidence on economics majors from three elite universities in the US that suggests attendance rates are slightly positively related to class size and are higher in upper level courses such as those we focus on in our analysis. There is an existing literature documenting the negative effect of absenteeism on academic achievement in university education [Romer 1993, Durden and Ellis 1995, Marburger 2001, Arulampalam *et al* 2007].

[24]As noted earlier, we focus on reporting effect sizes that are normalized by the within student standard deviation as that is the primary source of variation in class sizes we exploit. As the overall standard deviation in test scores (class sizes) is 6.67 (46.3), the effect size when the overall standard deviations are used instead of the within standard deviations is not much changed at $-.082$.

[25]Arcidiacono [2003] estimates a dynamic model of college and major choice to understand how students sort onto degree programs. While he reports there to be considerable variation in the returns to different majors, the majority of sorting occurs

One reason why class sizes vary is because some courses offered by a department are popular among students enrolled on degree programs in other departments, and their degree program regulations permit them to take these courses outside their own department. If student characteristics vary more between than within degree programs, then one concern is that in larger classes the diversity among students is mechanically larger. We therefore need to ensure that any class size estimates are not merely picking up any detrimental effects that student diversity may have on test scores. Such negative effects of student diversity may arise from, for example, faculty having difficulty in identifying the ability of the median student and tailoring their teaching appropriately.[26]

Column 4 addresses this concern by additionally controlling for a series of characteristics of students in the class, $H_c$. The result shows the class size effect to be very similar to the earlier estimates once the composition of the class is controlled for. This may be because the diversity of students in university classes has little effect on academic achievement, or that the positive and negative effects of having diverse peers approximately cancel out on average.

An important concern with these results is that departments may assign faculty to courses on the basis of class size. For example, departments may systematically assign stricter faculty to larger classes in order to keep discipline, and larger courses may in turn be more difficult. As in Lazear [2001], if class size and stricter discipline are substitutes, this confounds identification of any class size effect. To check for this Column 5 estimates (1) adding a complete series of faculty dummies, $\sum_j \mu_j f_{jc}$, as defined previously.[27]

The parameter of interest is $\widehat{\gamma} = -0.015$ which is significantly different from zero and implies an effect size of $-0.108$. Comparing the estimates with and without controlling for faculty dummies suggests these dummies are not much correlated with class size.[28] To check further for any evidence of departments non-randomly assigning faculty to class sizes, we re-estimated (1) without controlling for class size, $N_c$. We found no significant relationship between faculty quality in this specification, as measured by $\widehat{\mu}_j$, and the average class size the faculty member teaches – the line of best fit between the two has a slope coefficient not significantly different to zero. This suggests – (i) faculty are not assigned to class sizes on the basis of their underlying quality or style; (ii) faculty that are predominantly assigned to smaller classes do not systematically reward students with higher test scores, say because they can more easily observe students' effort throughout the academic year; (iii) students do not sort into classes on the basis of the quality or style of teaching faculty.

The reported effect sizes facilitate comparison between this setting of tertiary education vis-à-vis esti-

---

because of individual preferences for particular majors in college. Our data is consistent with such sorting by preferences – rather than class size – also explaining the choice of courses within any given program.

[26]On the other hand increased diversity among students may actually offset any negative effects of class size so that baseline estimates are upward biased. This would be the case for example if the most able or motivated students take courses outside their department and these types of student impose positive externalities onto their peers.

[27]Rothstein [2007] shows that such teacher effects are not identified in value added models if teachers are non randomly assigned to students over time, say because of ability tracking or parental pressure on school principals. In this setting such mechanisms are absent and we exploit the contemporaneous variation in class sizes and teaching faculty a student is exposed to in the same academic year. Each faculty member typically teaches on one or two courses each academic year. There is sufficient variation in the data to allow us to estimate $\mu_j$ for 550 teaching faculty in total.

[28]Further analysis shows that without controlling for the students fixed effects, $\alpha_i$, the teaching faculty dummies explain around 9% of the variation in student test scores. Hence characteristics of faculty members that have an equal effect on all students – such as their ability or motivational skills – are less important that student ability in explaining the variation in test scores as expected, although they remain more important than other measurable factors.

mates in the literature from primary or secondary school settings. However, an alternative measure of the magnitude of the effect in this specific setting is in terms of a coarser classification of exam performance as embodied in the following classmarks – an A-grade, corresponding to a final exam score of 70 and above, a B-grade (60 to 69), a C-grade (50 to 59), and a fail (49 or lower). This classification is relevant because UK employers make conditional offers to students on the basis of this classification rather than continuous test scores. Similarly, entry requirements for Ph.D. or professional qualifications courses are typically also based on this classification.

To assess the impact of class sizes on this metric of performance, we use a linear probability model analogous to (1) to estimate the effect of class size $N_c$ on the likelihood student $i$ obtains an A-grade on course $c$ in academic year $t$. We find this effect to be negative and significant. The magnitude of the coefficient implies that, evaluated from the mean, a one standard deviation increase in class size reduces the likelihood that the student obtains an A-grade by 1.1%. This is relative to a baseline probability of 13.4% across all courses. As offers of employment or acceptance onto doctoral study programs typically specify a particular classmark has to be obtained, it is possible that class size variations have short run effects on individual initial labour market outcomes or human capital accumulation.[29]

## 4.2   The Salience of Class Size

We present two pieces of evidence to assess whether the issue of class size is salient for students. First, we test whether class size is negatively related with students' satisfaction. To do so, we use students' reports of overall satisfaction with their course, as reported in student evaluations after the last lecture and before the final exam. This information is available only at the departmental-academic year level, not at the level of each course. With this caveat, we note that, as shown in Column 1 of Table 3, M.Sc. students are significantly less satisfied in departments with larger average class sizes, controlling for the overall number of students enrolled in the department. One concern is that average class size might be proxying for other departmental characteristics–for instance, the quality of the teaching faculty or administrative support. To allay this concerns, Column 2 regresses the reported satisfaction of undergraduate students on the average size of *postgraduate* classes. Intuitively, if the latter were to proxy for departmental quality, this should affect the satisfaction of all students, regardless of their level of studies. Reassuringly, Column 2 shows there is no correlation between the reported satisfaction of undergraduates and average class sizes at the M.Sc. level, so the results are unlikely to be picking up that student satisfaction differs systematically across departments. These results are in line with evidence from other university settings that students attitudes towards learning tends to be negatively affected by larger classes [Bolander 1973, Feldman 1984, McConnell and Sosin 1984].

Second, we test whether the magnitude of the class size effect is correlated with the number of courses

---

[29]In this linear probability model, 2.3% of the predicted values lie outside the zero-one interval, and the standard errors allow for the error term to be heteroskedastic. As an alternative approach, we also estimated a random effects ordered specification using maximum likelihood, where the random effect is the student effect and the likelihood for each observation is approximated by Gauss-Hermite quadrature using 15 points. It is not possible to also control for a complete series of lecturer dummies in this specification. This specification again shows a significant and negative class size effect. More precisely, individuals are significantly less likely to obtain a distinction (grade 70 and above) and significantly less likely to obtain a merit (grades 60-69), and are significantly more likely to obtain a pass (grades 50-59) in larger classes.

students can choose from. The rationale of this test is as follows. If students are aware that test scores are on average lower in larger classes, we expect them to choose smaller classes, other things equal. Whether it is feasible for students to substitute away from larger courses depends on the overall availability of courses. To measure this we exploit the fact that each of the 104 degree programs has its own rules determining the set of courses from which students have to take their core and elective courses. Degree programs vary in the number of available core courses – from one to thirteen, and the average program has 1.56 available core courses. Typically, students will have to choose one or two core courses from this list. The average program has 3.05 available electives, from which typically two or three can be chosen.[30] Denoting $L_i$ as the number of available courses to the student on her degree program, we then estimate the following panel data specification for core and elective course separately,

$$y_{ic} = \alpha_i + \gamma_0 N_c + \gamma_1 \left( N_c \times L_i \right) + \delta X_c + \lambda H_c + u_{ic}. \tag{2}$$

To the extent that students prefer smaller classes, we should find that any negative class size effects $(\widehat{\gamma}_0 < 0)$ are ameliorated by a larger choice set $(\widehat{\gamma}_1 > 0)$. As most faculty teach on both core and elective courses – rather than on many core courses for example – there is insufficient variation to identify the set of faculty dummies.[31]

Columns 3 and 4 of Table 3 estimate specification (2). The results show that the class size effect is less pronounced on cores when there are more core courses for students to choose from $(\widehat{\gamma}_0 < 0, \widehat{\gamma}_1 > 0)$. There is however no such ameliorating effect of greater choice on the class size effects among elective courses $(\widehat{\gamma}_0 < 0, \widehat{\gamma}_1 = 0)$, in line with the fact that students have far greater choice in their elective courses to begin with. Overall, the fact that greater choice helps to offset negative class size effects, suggests that students are aware of these effects and avoid them when possible. However, we note that departments are unable to provide students with sufficient course choice to fully offset the negative class size effects – the implied effect sizes remain negative and significant even for those students with the greatest choice of courses, as reported at the foot of Table 3.

## 4.3   Econometric Concerns

Identification of the causal effect of class size on students' performance is confounded by the possible presence of factors that are correlated to both variables. Decomposing the error term into two components, $u_{ic} = \varepsilon_{ic} + \varepsilon_c$, makes precise that class size and performance might be spuriously correlated because of omitted factors that vary at the student-course level or at the course level.

For instance, students may enrol into larger classes only for subjects for which they are intrinsically less motivated and seek smaller classes for subjects they care more about. To the extent that more motivation results in higher effort, this type of selection downwards biases any estimated class size effect. Note that, however, this and all other spurious mechanisms due to endogenous sorting by class size can only be at play if students have the option to choose among classes of different sizes for similar topics. If our previous

---

[30]Paglin and Rufolo [1990] and Arcidiacono *et al* [2007] present evidence from universities that students sort into courses on which they have a comparative advantage – which does not necessarily correlate to the size of the class.

[31]The correlation between the number of available core or elective courses ($L_i$) and class size ($N_{ct}$) is less than .06 in both cases. Hence it is possible to separately identify the effects of choice from class size.

estimates were spuriously generated by endogenous sorting we should then find a larger effect for students who have access to a larger range of substitutes. The fact that the effect is actually strongest when the availability of substitutes is low (Columns 3 and 4, Table 3) casts doubt on the relevance of this mechanism in this setting. This is also in line with the evidence on students' course choices in universities in North America presented in Paglin and Rufolo [1990] and Arcidiacono *et al* [2007].

Among course level omitted variables, a key candidate is course difficulty. In this setting, unlike in primary and secondary school, students have greater choice over which courses they enrol onto. To the extent that difficult courses attract more students, we would find a spurious negative class size effect, all else equal.[32] We address this by including a measure of the difficulty of the course directly into the baseline specification (1) – the share of students that are re-sitting it because they failed the same course in the previous academic year. Column 1 of Table A2 shows that the previous estimate is robust to including this measure of course difficulty.

To capture the effect of course level omitted variables that are time invariant, we exploit time variation to identify the effect of class size within the same course over academic years. Our first such specification is analogous to the baseline specification (1) except that we control for course fixed effects, $\alpha_c$, rather than student fixed effects, $\alpha_i$. With the inclusion of course fixed effects, we no longer control for a complete series of lecturer dummies because there is not much variation over time in the faculty that teach a given course, and nor are there multiple sections of the same course taught simultaneously. Robust standard errors are calculated. The result, in Column 2 of Table A2 shows there remains a negative relationship between class sizes and test scores, that is significant at the 5.1% significance level.

To address the concern that students might endogenously select their courses on the basis of class size, we re-estimate the baseline specification at the course-year level. This specification exploits only the time variation in class sizes within a course to estimate the parameter of interest, and allows the standard errors to be clustered by department-year to capture common shocks to all courses, such as those arising from a department's resource allocation. Column 3 of Table A2 shows that similar conclusions can be drawn to the earlier results if the data is collapsed to the course-year level.[33]

A related concern is that class size merely captures other omitted inputs that determine test scores, such as the quality of the lecture theatre, library textbooks, and computing facilities available to students. While we cannot directly control for such omitted factors, we note that *a priori*, most of these, should be linearly related to class size. To the extent that the relationship between class size and test scores is non-linear, omitted factors alone are unlikely to explain the findings. The next Section analyses this issue in detail.

---

[32]In a prestigious university as the one in this study, students are selected from the right tail of the ability distribution and may therefore be keen to enrol onto the most challenging courses. Arcidiacono [2003] finds, using data from the NLS72, that students who perform better than expected on their major are more likely to stay on the same major or switch to a more difficult major. On the other hand if students are attracted to courses that are perceived to be easier, this leads to a positive class size effect so that we underestimate the true effect of class sizes on test scores.

[33]More precisely, Column 3 estimates the following panel data specification at the course-year level,

$$\overline{y}_{ct} = \alpha_c + \gamma N_{ct} + \delta X_{ct} + \lambda H_{ct} + \sum_j \mu_j f_{jct} + u_{ct},$$

where $\overline{y}_{ct}$ is the average test score of all students on course $c$ in academic year $t$, $\alpha_c$ is a course fixed effect, and all other controls are as previously defined, and the error term is clustered by department-academic year.

# 5    Heterogeneous Class Size Effects

The mechanisms that link class size to achievement can operate differently at different levels of class size, or may have heterogeneous effects across students. To shed light on why class size matters for achievement, we now exploit the full richness of the data to explore these two forms of heterogeneous class size effect that have not been previously documented within university education.

## 5.1    *Non-Linear Effects*

We first assess whether the class size effect is non-linear using the following panel data specification,

$$y_{ic} = \alpha_i + \sum_{q=2}^{5} \gamma_q D_{qc} + \delta X_c + \lambda H_c + \sum_j \mu_j f_{jc} + u_{ic}, \tag{3}$$

where $D_{qc}$ is equal to one if the class size is in $q$th quintile of class size distribution, and zero otherwise. All other controls are as previously defined and we continue to cluster $u_{ic}$ by course-academic year. An important feature of this empirical setting is that the same student is exposed to class sizes of very different size. As Figure 3B shows, the median student has a class size gap of 56, and more than 20% of students have a class size gap of at least 100, therefore spanning all five quintiles of the class size distribution. An implication is that it is feasible to include student fixed effects $\alpha_i$ in (3) and estimate the $\gamma_q$ coefficients using class size variation within the same student.[34]

Table 4 presents the results. To begin with, Column 1 estimates (3) without controlling for the series of faculty dummies. The result shows there to be a non-linear effect of class size on student test scores. More precisely, there is a negative and significant class size effect moving from the first quintile, which corresponds to class sizes of 1 to 19, to the second quintile (20-33), so $\widehat{\gamma}_2 < 0$, and then another negative and significant class size effect moving from the second to the third quintile (34-55), so $\widehat{\gamma}_3 < 0$. Importantly, there is no additional class size effect moving from the third to the fourth quintile (56-103), so $\widehat{\gamma}_4 = \widehat{\gamma}_3$. Finally, there is an additional negative class size effect moving from the fourth to the fifth quintile (104-211), so $\widehat{\gamma}_5 < 0$. At the foot of the table we report p-values on two-sided t-tests of $\gamma_q = \gamma_{q+1}$. These confirm that the class size effects increase in absolute magnitude moving from the second to the third quintile, and from the fourth to the fifth quintile, yet there is no detrimental class size effect moving from the third to the fourth quintiles.[35]

Moving from the first to the second quintile, $\widehat{\gamma}_2$ implies an effect size of $\frac{\widehat{\gamma}_2}{sd(y_c)} \approx -0.107$ where the standard deviation in test scores is within student. The implied effect size is around four times larger (.460) moving from quintile 1 to quintile 5 – using the midpoint values in each quintiles, this would correspond to increasing class size from 10 to 158. Normalising this effect size by the overall standard deviation in test scores, this corresponds to an effect size of $-0.304$, equivalent to moving a student from the median to the 25th percentile in cumulative distribution of test scores. This is sufficiently large to

---

[34]The quintile the class size is in is defined relative to the distribution of class sizes over the 1775 courses by academic year that are observed in the data.

[35]This non-linear effect was also found using alternative methods including those that impose more parametric structure such as controlling for a cubic polynomial in class size, as well as using semi-parametric estimation techniques.

move an individual across classmarks, and so can have long run effects on individuals over their life cycle as it affects their labour market outcomes upon graduation, and their opportunities to continue into doctoral programs or to study for professional qualifications.

The implications for university policies for whether class sizes should be limited or capped are rather stark. In terms of student achievement, there are clear gains to be had from reducing class sizes in two cases. First, if class sizes are sufficiently small to begin with and can be reduced from the second to the first quintile of class sizes which approximately translates to reducing class sizes from above the mid 30s to below 20. Second, if class sizes are very large to begin with and can be capped or reduced from the fifth to the fourth quintile. This approximately translates to not allowing class sizes to reach limits above 100. However, there appears to be little or no benefit – as measured by student test scores – of reducing class sizes over a wide range of intermediate class sizes, approximately corresponding to between the mid 30s to around 100.[36]

These basic conclusions are not much changed when unobserved heterogeneity across faculty members is controlled for, as reported in Column 2. The magnitude of the $\gamma_q$ coefficients are very similar across Columns 1 and 2. The pattern of $\gamma_q$ coefficients from Columns 1 and 2 are shown in Figure 4. The inclusion of a complete series of faculty dummies into (3) leaves unchanged the basic implication that class size effects are non-linear. Moreover, the results make clear that, in line with the results in Table 2, there is no systematic assignment of better or worse faculty to classes of different size, at any quintile of class size.

One concern is these non-linear effects may be picking up differential effects across core and elective courses – as shown in Table 1 core courses tend to be larger than electives by virtue of the fact that they are often compulsory for students to attend. To address this, Columns 3a and 3b estimate (3) for core and elective courses separately. The results show the negative class size effect in the smallest courses – moving from the first to the second quintile of class sizes – is significantly more pronounced in cores than in electives. The negative class size effect in the largest courses – moving from the fourth to the fifth quintile of class sizes – is also more pronounced in core courses. For both types of course, the evidence suggests there exists a wide range of intermediate class sizes over which small changes in class sizes would have little impact on student test scores. For core courses this range extends from the second to the fourth quintile, while for electives this range covers the third and fourth quintiles.[37]

In line with our earlier findings, the comparison of core and electives indicates that the negative effect of class size is more pronounced when students have less choice – namely, for core courses. This is again suggestive of the fact that students avoid larger classes when they can, in line with the hypothesis that they are aware of the negative effect of class size.

The finding that class size effects are non-linear also helps address concerns that our estimates capture

---

[36]This finding is in line with the results in Machado and Vera Hernandez [2008] on class size effects for first year undergraduates in a leading Spanish university. They observe class sizes in the 50-70 range, and find a zero effect size on test scores. Similarly, Kokkelenberg *et al* [2007] report non linear class size effects using data from undergraduates in a northeastern public university – they report large negative class size effects up to 40 and smaller negative effects thereafter. Finally, Duflo *et al* [2007] report results from a randomized experiment that reduced class sizes from 80 to 46 in Kenyan primary schools and found no effect on test scores.

[37]Again because faculty tend to teach core and elective courses there is insufficient variation to also identify the complete series of faculty dummies within only core or elective courses. To facilitate comparison with the earlier results, we maintain the same class sizes in each quintile for these specifications.

other factors that vary with class size including – (i) unobserved inputs such as the availability of computers and library textbooks; (ii) the difficulty of the course; (iii) benevolence in grading. This is because each of these factors would *a priori* be expected to vary linearly with class size, and also to typically vary in the same way across core and elective courses.[38,39]

The result that there are two distinct ranges over which negative and significant class size effects exist suggests there are at least two mechanisms at play. The existence of a variety of possible mechanisms also makes it more likely that, on the margin, reducing class size may have differential effects across students depending on the students' ability, as well as on the initial level of class size. To shed light on this, we now explore in more detail whether there are heterogeneous class size effects across students.

## 5.2  *Quantile Regression Estimates*

We use quantile regression to estimate whether the previously documented class size effects on test scores differ across students. More formally, we estimate the following specification to understand how the conditional distribution of the test score of student $i$ on course $c$ in year $t$, $y_{ic}$, is affected by class size $N_c$ at each quantile $\theta \in [0, 1]$ of the distribution,[40]

$$Quant_\theta(y_{ic}|.) = \gamma_\theta N_c + \delta_\theta X_c + \lambda_\theta H_c, \tag{4}$$

where $X_c$, and $H_c$ are as previously defined in Section 4. We first estimate (4) using the full range of class sizes. Figure 5A plots the implied effect size from the $\widehat{\gamma}_\theta$ coefficients at each quantile $\theta$, the associated 95% confidence interval, and the corresponding OLS estimate from (1) as a point of comparison. This shows – (i) the effect size is negative at each quantile $\theta$; (ii) the effect size is larger in absolute value at higher quantiles. This suggests the conditional distribution of test scores becomes more compressed with increases in class size, and that this compression occurs because those students at right tail of the conditional distribution of test scores – whom we refer to as 'high ability' students – are more affected by increases in class size.[41]

To understand whether the distributional effects of class sizes vary with the size of the class, we use

---

[38] In this university there is a specific funding algorithm in place so that the resources available to the library to cater for the students on any given course are proportional to the number of students enrolled on the course. In addition, the physical characteristics of lecture theatres are such that the capacity of lecture theatres varies continuously with the largest lecture theatres having a capacity of over 200.

[39] These results help address the concern that in smaller class sizes faculty may get to know students better and be more appreciative of the effort students exert throughout the academic year. This may then be reflected in higher exam marks in smaller courses leading to a spurious negative class size effect. Although we cannot rule this out altogether, the fact we find negative class size effects to exist in larger classes suggests this explanation alone does not explain the full pattern of class size effects found. Moreover, in the smallest class sizes we might expect this effect to operate equally in core and elective courses – contrary to the pattern of coefficients in Columns 3a and 3b. Finally, we reiterate that exam scripts are anonymous and typically are not marked solely by the faculty member teaching the course.

[40] This approach is particularly applicable in this context because the dependent variable, student's test score, is measured without error.

[41] Arulampalam *et al* [2007] also use quantile regression on administrative data on second year undergraduate economics students at a UK university to explore the distributional consequences of absenteeism on test scores. They also find differential effects of absenteeism at the top end of the conditional distribution of test scores, with the most able students being most affected by absenteeism. In contrast, Duflo *et al* [2007] find high ability students in primary school to be unaffected by class size reductions.

the following specification to estimate whether class size effects are heterogeneous across students moving from a class size in quintile $q$ to one in quintile $q'$,

$$Quant_\theta(y_{ic}|.) = \sum_{q=2}^{5} \gamma_{\theta q} D_{qc} + \delta_\theta X_c + \lambda_\theta H_c, \tag{5}$$

where $D_{qc}$ is set equal to one if the class size is in $q$th quintile of class size distribution, and zero otherwise, and all other controls are as previously defined. To estimate (5) we then only use observations from class sizes in any two adjacent quintiles $q$ and $q'$. Figure 5B then plots $\widehat{\gamma}_{\theta q}$ at each quantile $\theta$ – the effect on the conditional distribution of test scores at each quantile $\theta$ of moving from quintile $q-1$ to quintile $q$ in the class size distribution, and the associated 95% confidence interval.

Three points are of note. First, moving from the first to the second quintile of class size, there is a uniformly negative effect on test scores at all quantiles. Changes in student or faculty behaviour that induce these class size effects impact equally the test scores of high and low ability students. This would be consistent, for example, with either all students being less attentive, on the margin, as class sizes increase over this range, or the fact that faculty are less able to identify the needs of the median student say, having similarly detrimental effects on the learning of all students.

Second, there are no heterogeneous class size effects over a wide range of intermediate class sizes corresponding to the third and fourth quintiles of class size, in line with the results in Table 4.[42]

Third, moving from the fourth to the fifth quintile of class size, there is a negative class size effect for most quantiles $\theta$. However, the magnitude of the effect is significantly larger in higher quantiles – the effect on student test scores in the lowest quantiles is close to zero and precisely estimated, and the implied effect among the highest quantiles is slightly larger than $-2$. This suggests changes in student or faculty behaviour that induce such class size effects appear to impact the test scores of high ability students to a greater extent, and highlights there is an important complementarity between student ability and class sizes that exists only in the largest class sizes.[43]

# 6   Mechanisms

## 6.1   *Teachers' behaviour*

One possible mechanism linking class size to students achievement acts through changes in teachers' behaviour. For instance, teachers might devote less effort to larger classes because they are unable to assess

---

[42]For expositional ease we do not show the estimates of moving from quintile 2 to quintile 3. In line with the results in Table 4, these are slightly more negative than moving from quintile 1 to quintile 2 at each $\theta$.

[43]One concern is that the results may merely reflect that more able students can afford to slack and still remain within the same overall classmark. For example the most able students may anticipate a test score well above 70 and so are more able to slack and remain confident of achieving a test score of at least 70 and therefore an A-grade on the course overall. This effect however should only depend on the student's ability and not on the size of the class she is enrolled in. A similar concern would be that in larger courses, a greater number of graders are required to mark exam scripts. Hence it is more likely that such graders are unfamiliar with the course materials and therefore find it harder to identify the best and worse students. If this were the case then we would expect a compression of exam grades at both the left and right hand tails of the conditional test score distribution, and an effect on test scores moving from the third to fourth quintiles of class size. Neither of these implications is supported by the data.

the students' needs. The premise for our analysis is the earlier finding that, while their assignment is uncorrelated to class size, teachers do affect students' performance. As noted above, the faculty dummies explain around 9% of the variation in student test scores. Moreover, 74% of the faculty dummies $\widehat{\mu}_j$ , estimated in Column 5, Table 2, are significantly different from zero.[44]

A closer inspection of the estimated $\widehat{\mu}_j$'s reveals enormous heterogeneity in the effect of different teaching faculty. The difference between the 10th and the 90th percentiles of $\widehat{\mu}_j$ is 5.52, that is one third of the interdecile range of the unconditional distribution of test scores. Moreover, the standard deviation of the $\widehat{\mu}_j$'s is 2.59. This suggests there can be large gains in student learning, as measured by test scores, moving one standard deviation in the distribution of 'faculty quality', $\widehat{\mu}_j$. As in Rivkin *et al*'s [2005] study of the role of teacher quality in primary school settings, altering faculty inputs in university settings has quantitatively large effects. Whether, evaluated from the mean, this has greater effects on test scores than marginal adjustments in class sizes is unclear because both are presumably extremely costly adjustments to make.[45]

To understand whether and how the same faculty member alters her behaviour when exposed to larger class sizes we exploit the fact that in this empirical setting we observe within faculty variation in class sizes taught. Denoting the class size taught by professor $j$ on course $c$ as $N_{jc}$, we then estimate the following specification,

$$\overline{y}_{jc} = \sum_j \mu_j f_{jc} + \sum_j \gamma_j [N_{jc} - \overline{N}_j] f_{jc} + \delta X_c + \lambda H_c + u_{jc}, \tag{6}$$

where $\overline{y}_{jc}$ is the average test score of students on faculty member $j$'s course $c$, and $\overline{N}_j$ is the average class size taught by faculty member $j$ over the sample period – academic years 1999/00-2003/04. $f_{jc}$ is a dummy equal to one if faculty member $j$ teacher course $c$ and is zero otherwise, and $X_c$ and $H_c$ are as previously defined, and we allow the error term to be heteroskedastic.

The coefficient of interest $\gamma_j$ measures how the average test scores of students of faculty member $j$ are affected by that faculty member teaching a class that is larger than they experience on average. The null hypothesis is that $\gamma_j = 0$ so faculty do not alter their behaviour in classes of different size. This may be due to most of the costs to faculty of teaching being fixed, so that how faculty prepare and present lectures is independent of the numbers of students. On the other hand, $\gamma_j$ may be positive if faculty devote more effort to class preparation when more students stand to benefit, for example. Finally, $\gamma_j$ may be negative if in larger than average classes, faculty are unable to tailor their teaching to appeal to the median student, they receive less feedback from students on how they can improve the course to meet the needs of students, or it is simply harder for them to monitor whether students are paying attention in the lecture.

Estimating (6) reveals heterogeneous responses to class sizes by faculty that are consistent with all

---

[44]This is in line with test scores not being curved because the identity of the teaching faculty matters. Moreover, the $\widehat{\mu}_j$'s also do not significantly differ by faculty rank, in line with faculty not solely grading their own course. There are therefore few opportunities for faculty to strategically manipulate test scores to attract students to their course, which more junior faculty, or faculty that teacher smaller courses, may otherwise feel pressure to do.

[45]In the literature on the effects of teacher quality on test scores in primary and secondary education settings, a number of studies find that a one standard deviation increase in teacher quality raises student scores by around .10 to .25 of a standard deviation [Aaranson *et al* 2007, Rockoff 2004, Rivkin *et al* 2005, Kane *et al* 2006], although these approaches have been criticized by Rothstein [2007]. In university settings, Hoffman and Oreopoulos [2006], using administrative data from a Canadian university, find a one standard deviation increase in teaching faculty quality leads to an increase in grades by .5 percentage points, which is qualitatively smaller than the impact in primary and secondary settings.

three mechanisms. However, we note that for 76% of faculty, $\widehat{\gamma}_j$ is not significantly different from zero – hence the data support the idea that fixed costs of preparation dominate other possible effects, so that faculty behaviour is unaffected by the number of students on the course. Hence any class size effects do not predominantly appear to originate from changes in faculty behaviour across classes of different size.[46]

## 6.2    Students' behaviour

As a final step, we exploit the observed heterogeneity in students' characteristics to test the relevance of two mechanisms through which class size affects test scores. First we test whether class size reduces test scores because it exacerbates informational constraints. Second we test whether class sizes reduces test scores because it reduces the provision of other complementary inputs, such as access to books or computer equipment. In each case we estimate the following panel data specification,

$$y_{ic} = \alpha_i + \sum_q \gamma_q^0 D_{qc}(1 - Z_i) + \sum_q \gamma_q^1 N_{qc}(Z_i) + \delta X_c + \lambda H_c + \sum_j \mu_j f_{jc} + u_{ic}, \tag{7}$$

where $Z_i$ is a dummy variable defined below, and all other variables are as previously defined. We continue to cluster the error term by course-academic year. The parameters of interest are $\gamma_q^0$ and $\gamma_q^1$ at each quintile $q$ – a comparison of these coefficients is then informative of whether the class size effects are more pronounced for some students rather than others.

We use two proxies of the severity of informational constraints students face – (i) $Z_i = 1$ for British students and zero otherwise; (ii) $Z_i = 1$ for students who obtained their undergraduate degree from this same institution and zero otherwise. The rationale is that students who have already experienced the UK university system are more familiar with the examination style, the extent to which students are expected to work on their own, and norms regarding how approachable faculty are during their office hours for example. In addition, students who are familiar with this institution are more likely to have personal contact with former M.Sc. students and therefore be more informed about the choice of courses on different degree programs, the difficulty of courses, and the quality of faculty.[47]

The results are presented in Columns 1 and 2 of Table 5. All students – as defined along both dimensions – experience similar non-linear class size effects. As reported at the foot of Table 5, $\widehat{\gamma}_q^0$ is never significantly different from $\widehat{\gamma}_q^1$ at any quintile $q$ nor for any individual characteristic $Z_i$. We thus find no evidence that class size affects achievement through informational constraints.

To assess whether larger classes offer fewer complementary inputs and this reduces test scores, we identify the students who are more likely to be able to replace these complementary inputs. Intuitively,

---

[46]The remaining 24% of faculty have $\widehat{\gamma}_j \neq 0$ with almost equal numbers having $\widehat{\gamma}_j > 0$ and $\widehat{\gamma}_j < 0$. Although equal numbers of faculty have positive and negative interaction effects, the distribution is right skewed. This suggests among faculty whose behavior significantly changes with class size, there is a slightly more pronounced effect for student achievement to fall in larger than average class sizes.

[47]Students that were undergraduates at the same institution and British students form two distinct groups, each of which accounts for 17% of the sample. This is because there are 153 nationalities represented in the sample. Indeed, there are only 265 sample students that are both British and former undergraduates, 1627 British students that were not former undergraduates, and 1601 former undergraduates that are not British. On average, former undergraduates at the university do not take courses with significantly smaller class sizes than other students despite being more likely to be aware of class sizes. Nor do they have significantly different test scores.

wealthier students may be able to purchase inputs, such as textbooks or personal tutors, that offset some forms of negative class size effect. The administrative records contain the postcode of residence for each student. We use this information for the subset of students that reside in private residences to form a proxy for their family income based on the value of house price sales in their postcode. 68% of sample students reside in private residences as opposed to university residences. We use the average price of sold flats/maisonettes in the three digit postcode as these are correlated to rental rates for students' accommodation. We then define a wealthy student ($Z_i = 1$) to be one that resides in a postcode with higher than median values of house price sales in the academic year in which $i$ attends the university, and $Z_i = 0$ otherwise.[48]

The findings in Column 3 show that the pattern and magnitude of the class size effect is identical for the two groups of student. This helps rule out the documented class size effects could be offset by students' ability to purchase replacement inputs in the marketplace. This leaves open the possibility that the class size effects stem from some non-market mechanism, such as changes in behaviour of students or the provision of feedback, or references, to students from faculty in class sizes of different size.[49]

While more research is clearly required to pin down this mechanism in university settings, we note that in a companion paper using the same administrative data on this same set of students [Bandiera *et al* 2008], we find robust evidence that – (i) the provision of feedback to students about their past academic performance has positive and significant effects on their future academic performance, as measured by test scores; (ii) the effect of feedback on test scores is indeed significantly larger on more able students, consistent with the quantile regression estimates in Section 5.2.

# 7    Discussion

It is widely perceived among the general public and university administrators that class sizes are an important input into the educational production function. We contribute to this debate by providing some of the first estimates of the impact of class size on students' academic achievement, as measured by their end of year exam test scores. Against a backdrop of steadily increasing enrolment rates into universities in the UK – as in much of the OECD – our analysis has the following policy implications.

First, there is robust evidence of a negative class size effect – on average, larger classes reduce students academic achievement as measured by test scores. However the magnitude of this effect is generally smaller than has been documented to be the case at other tiers of the education system. We also document that there exists a wide range of class sizes over which, on the margin, targeting resources to reduce class sizes will have little impact on test scores. However, reducing the very largest class sizes – namely those over

---

[48]House price information is available at the three digit zip code level – zip codes in the UK typically have six digits in total. The first three digits correspond to areas that are far smaller than US counties. There are 282 unique three digit zip codes among the sample students. House price information was obtained from the Land Registry at http://www.landregistry.gov.uk/.

[49]The most important form of student's change in behaviour is obviously the effort and time devoted to studying, something on which we have no information. On this point, Stinebrickner and Stinebrickner [2008] estimate that an increase in studying of one hour per day causes first semester GPAs to rise by 0.36, which is more than half of the GPA standard deviation. In our analysis we cannot establish whether, or under which circumstances, study effort and class size are complements or substitutes. However, any change in students' behaviour which is induced purely by class size should be considered a class size effect.

100 – will significantly raise test scores, especially for the most able students.

Our analysis also sheds some light on the impact of altering faculty inputs. As in Rivkin *et al* [2005]'s analysis of primary schools, we find that altering faculty inputs has quantitatively large effects on test scores. However without further information on the cost of such input adjustments, it is impossible to say whether targeting resources to either hire better faculty or to reduce class sizes – in the relevant range of class sizes – is the most cost effective means for universities to raise academic achievement. This is certainly worthy of further research given the continued rise in tuition fees in UK universities, at least among the most popular universities, that has increased the resources available to higher education institutions.[50]

Further research is also required to better understand the underlying mechanisms that drive the effect of class size. Our analysis rules out that the class size effects are due solely to the non-random assignment of faculty to class size, sorting by students onto courses on the basis of class size, omitted inputs, the difficulty of courses, or grading policies. The evidence also shows that these class size effects are not mitigated for students with greater knowledge of the UK university system, this university in particular, or with greater family wealth. This casts doubt on the relevance of mechanisms that work through the students' knowledge of the institution and those that work through input congestion, as wealthier students should be better able to procure these inputs privately. The fact that the effect of class size is non-linear indicates that different mechanisms operate over different ranges. In the smallest class size ranges, the ability of faculty to identify the needs of the median student might quickly diminish with incremental changes in class size. However, once the class is sufficiently large this mechanism is no longer relevant. In the largest class size ranges, the ability of students to obtain feedback from faculty on whether their studying style is appropriate for the course might decline with increases in class sizes. This would be in line with the evidence from a companion paper [Bandiera *et al* 2008], which shows that the effect of providing feedback is larger on the most able students.

While our findings are based on the analysis of the records of one particular institution, three pieces of evidence shed light on the generalisability of the policy implications. First, we note that the institution we study has experienced an increase in enrollment which is in line with the national trend, and its student-to-staff ratio (13) is identical to the average figure for the top 20 universities in the country (12.9) [O'Leary *et al* 2003]. Second, we find that the non-linear class effects does not differ by the student's nationality, familiarity with the university, and a proxy for their income. This suggest the results would hold in other universities that draw their students from a different pool of nationalities, or from different parts of the income distribution. Third, we find that high ability students are more affected by larger class sizes. If this is driven by the ability of the students relative to their peers', we expect similar effects to arise in other settings, as there would always exists a group if relatively more able students in other university settings who might be disadvantaged by larger class sizes. If, in contrast, this is driven by absolute ability, we expect the effect of class size to be homogeneous in universities that attract students of lower absolute ability. Certainly if the class size effect stems from students being unable to receive feedback from faculty, then similar effects can be expect to be found in other universities.

Finally, we note that to the extent that the demand for higher education is countercyclical, the need

---

[50]Other analyses of specific supply side policies related to higher education include the effects of tuition fees [Fortin 2006] and scholarships and mentoring services [Angrist *et al* 2007].

to understand the consequences of larger class sizes may be especially urgent with the downturn in the global economy. To this end, there may be important complementarities to explore in future research between the microeconomic literature this paper contributes to on the organisation of universities, and the hitherto separate macroeconomic literature on the effects of adverse economic conditions at the time of labour market entry over the life cycle [Gibbons and Waldman 2004, Oyer 2008].

# 8    Appendix: Input Adjustment

We present evidence that departments cannot adjust on all margins in response to aggregate changes in student enrolment. This opens up the possibility that changes in enrolment partially feed through into undesired changes in class size which may in turn affect student achievement as measured by end of year test scores. We focus first on inputs in the form of numbers of teaching faculty. We also distinguish between two types of student that can enrol onto courses offered by a given department $d$ – either students can be registered on a degree program offered by department $d$ itself, or students can be enrolled on related programs in other departments $d'$ and their degree program regulations allow them to take courses in department $d$. We therefore estimate the following specification for inputs into department $d$ in academic year $t$ $(y_{dt})$,

$$y_{dt} = \alpha_d + \beta_0 E_{dt} + \beta_1 R_{dt} + u_{dt}, \tag{8}$$

where $y_{dt}$ refers to the number of teaching faculty, $E_{dt}$ is the number of students enrolled in department $d$, and $R_{dt}$ is the number of students enrolled in related departments. Note that $R_{dt}$ is $dt$ specific so that each department has a series of bilateral agreements with a subset of other departments over whether students registered on programs offered by department $d'$ are permitted to take courses organised and run by department $d$, and these agreements can change over time. By controlling for department fixed effects $\alpha_d$ we only exploit variation in year to year student enrolments and therefore shed light on whether and how student enrolments correlate to departmental inputs such as numbers of teaching faculty. We allow the error term $u_{dt}$ to follow an AR(1) process where the autocorrelation coefficient is restricted to be the same across departments.

Table A1 presents the results. Column 1 shows that as the number of students enrolled in the department increases, the number of teaching faculty also significantly increases $(\widehat{\beta}_0 > 0)$. The magnitude of the coefficient implies if the number of students enrolled in department $d$ were to increase by 17.7, this would be associated with there being one more faculty member teaching, as reported at the foot of Column 1. In contrast, there is no correlation between the number of teaching faculty and the number of students enrolled in related departments and so who could potentially attend courses offered by department $d$ – we find $\widehat{\beta}_1 \approx 0$ and $\widehat{\beta}_0$ is significantly different to $\widehat{\beta}_1$.

Columns 2 to 4 break this result down by faculty rank. We see that as more students enrol in department $d$ the number of full and other professors that teach, significantly increases. As expected, the increase in enrolment of students registered with department $d$ associated with one more full professor teaching (37.9) is larger than the increase in enrolment associated with either an associate or assistant professor teaching (26.4). In contrast, we again see that the number of teaching faculty is uncorrelated with students that

enrol in related departments and so can potentially attend courses offered by department $d$. Finally, Column 4 shows that enrolment for neither source – in department $d$ or related departments $d'$ – affects the number of non-professors that teach.[51]

Taken together these results show that although departmental inputs in terms of teaching faculty do partially adjust to student enrolments, they do so only in response to students that enrol into department $d$ itself, and are unrelated to those students that can enrol onto programs in department $d'$ are are eligible to enrol onto courses offered by department $d$. In other words the resources departments have to finance teaching faculty appear to be related to the number of students enrolled in the department ($E_{dt}$), not directly related to the number of students actually taught.

However, both sources of student – namely those enrolled into department $d$ and those enrolled in related departments – influence class sizes in courses offered by department $d$. More formally, Column 5a shows that student enrolments in the own and related departments are both significantly associated with larger class sizes in courses organised by department $d$. The coefficients imply that if 8.95 more students enrol in department $d$ then class size will increase by one on the average course offered by department $d$, and class sizes will on average increase by one if 40 more students enrol in related departments. Column 5b shows these effects remain controlling for the number of courses and programmes offered by department $d$. Hence even if departments respond to increased enrolment by adjusting along these margins, these adjustments do not appear to be sufficient to prevent class sizes from increasing overall.[52]

These results show that departments can adjust inputs to a greater extent in response to changes in their own student enrolment, than in response to changes in enrolment in related departments ($|\beta_0| \geqslant |\beta_1|$). However, class sizes are positively correlated with both sources of student enrolment. Taken together this implies that student enrolments in department $d'$ impose a negative externality onto class sizes in related department $d$. This negative externality stems from the fact that – (i) departments cannot deny students from related departments enrolling onto their courses; (ii) the resources departments command for organising and running their courses relate to the number of students enrolled in the department, not the numbers of students actually taught.[53]

---

[51] These figures are in line with anecdotal evidence given to us by heads of department suggesting that if around 25 more students enrol onto programs in the department, the department is often able to negotiate additional resources from the university to hire one more faculty member to teach.

[52] Note that on average a department offers 16 courses, of which students take around 4 or 5. Hence we expect $\beta_0$ to be far smaller than one in the class size regressions in Columns 5a and 5b. Second, although students in related departments can potentially take courses offered by the department, many of them will choose not to do so. Hence we also expect $\beta_1$ to be far smaller than $\beta_0$ in absolute value.

[53] These results beg the question why departments are allowed to impose this negative externality onto each other? Although this lies outside the scope of this paper, we speculate it may relate to there being potentially a very large number of bilateral agreements that would need to be considered for these externalities to be internalized – in this university there are 24 departments and so potentially up to ($\frac{1}{2}(24 \times 23)$) bilateral agreements across departments. Moreover, the flows of students from department $d$ to $d'$ need not be symmetric to flows from $d'$ to $d$. In addition, it is far more straightforward from an accounting perspective to reward department on the basis of students enrolled on programs offered by the department, rather than based on the number of students actually taught.

# References

[1] AARONSON.D, L.BARROW, AND W.SANDER (2007) "Teachers and Student Achievement in the Chicago Public High Schools", *Journal of Labor Economics* 24: 95-135.

[2] AGHION.P, L.BOUSTAN, C.HOXBY, J.VANDENBUSSCHE (2007) Exploiting States' Mistakes to Identify the Causal Impact of Higher Education on Growth, mimeo, Harvard University.

[3] ANGRIST.J AND V.LAVY (1999) "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement", *Quarterly Journal of Economics* 114: 533-75.

[4] ANGRIST.J, P.LANG, AND P.OREOPOULOS (2007) Incentives and Services for College Achievement: Evidence from a Randomized Trial. mimeo MIT.

[5] ARCIDIACONO.P (2004) "Ability Sorting and the Returns to College Major", *Journal of Econometrics* 121: 343-75.

[6] ARCIDIACONO.P, G.FOSTER, N.GOODPASTER AND J.KINSLER (2007) Estimating Spillovers Using Panel Data, With an Application to the Classroom, mimeo, Duke University.

[7] ARCIDIACONO.P AND S.NICHOLSON (2005) "Peer Effects in Medical School", *Journal of Public Economics* 89: 327-50.

[8] ARULAMPALAM.W, R.A.NAYLOR AND J.SMITH (2007) Am I Missing Something? The Effects of Absence From Class on Student Performance, mimeo, University of Warwick.

[9] BANDIERA.O, V.LARCINESE AND I.RASUL (2008) Blissful Ignorance? Evidence From a Natural Experiment on The Effect of Individual Feedback on Performance, mimeo LSE.

[10] BLATCHFORD.P, P.BASSETT AND P.BROWN (2005) "Teachers' and Pupils' Behaviour in Large and Small Classes: A Systematic Observation Study of Pupils Aged 10/11 Years", *Journal of Educational Psychology* 97: 454-67.

[11] BOLANDER.S.F (1973) "Class Size and Levels of Student Motivation", *Journal of Experimental Design* 42: 12-8.

[12] BOYD-ZAHARIAS.J AND H.PATE-BAIN (2000) "Early and New Findings from Tennessee's Project STAR", in M.C.Wang and J.D.Finn (eds.) *How Small Classes Help Teachers Do Their Best*, Philadelphia, PA: Laboratory for Student Success at Temple University Center for Research in Human Development and Education.

[13] DUFLO.E, P.DUPAS AND M.KREMER (2007) Peer Effects, Pupil Teacher Ratios, and Teacher Incentives: Evidence from a Randomized Evaluation in Kenya, mimeo, MIT.

[14] DURDEN.G.C AND L.V.ELLIS (1995) "The Effects of Attendance on Student Learning in Principles of Economics", *American Economic Review* 85: 343-6.

[15] FELDMAN.K.A (1984) "Class Size and College Students' Evaluations of Teachers and Courses: A Closer Look", *Research in Higher Education* 21: 45-91.

[16] FINN.J.D, S.B.GERBER, C.M.ACHILLES, AND J.BOYD-ZAHARIAS (2001) "The Enduring Effects of Small Classes", *Teachers College Record* 103: 145-83.

[17] FORTIN.N.M (2006) "Higher Education Policies and the College Premium: Cross-State Evidence from the 1990s", *American Economic Review* 96: 959-87.

[18] GIBBONS.R AND M.WALDMAN (2004) "Task-Specific Human Capital", *American Economic Review* 94: 203-7.

[19] GLAESER.E.L, H.D.KALLAL, J.SCHEINKMAN, AND A.SHLEIFER (1992) "Growth in cities", *Journal of Political Economy* 100: 1126-51.

[20] HANUSHEK.E.A (2003) "The Failure of Input Based Schooling Policies", *Economic Journal* 113: F64-98.

[21] HOFFMAN.F AND P.OREOPOULOS (2006) Professor Qualities and Student Achievement, mimeo, University of Toronto.

[22] HOXBY.C (2000) "The Effects of Class Size on Student Achievement: New Evidence from Natural Population Variation", *Quarterly Journal of Economics* 116: 1239-86.

[23] JOHNSTON.J.M (1990) Effects of Class Size on Classroom Processes and Teacher Behaviors in Kindergarten Through Third Grade, ERIC Document Reproduction Service No. ED321848.

[24] KANE.T.J, J.E.ROCKOFF AND D.O.STAIGER (2006) What Does Certification Tell Us About Teacher Effectiveness? Evidence From New York City, NBER Working Paper 12155.

[25] KOKKELENBERG.E.C, M.DILLON AND S.CHRISTY (2007), "The Effects of Class Size on Student Grades at a Public University", forthcoming *Economics of Education Review*.

[26] KRUEGER.A (1999) "Experimental Estimates of Education Production Functions", *Quarterly Journal of Economics* 114: 497-532.

[27] LAZEAR.E (2001) "Educational Production", *Quarterly Journal of Economics* 116: 777-803.

[28] MACHADO.M.P AND M.VERA HERNANDEZ (2008) Does Class Size Affect the Academic Performance of First Year College Students?, mimeo University College London.

[29] MARBURGER.D.R (2001) "Absenteeism and Undergraduate Exam Performance", *Journal of Economic Education* 32: 99-109.

[30] MARTINS.P AND I.WALKER (2006) Student Achievement and University Classes: Effects of Attendance, Size, Peers, and Teachers, mimeo Warwick University.

[31] MCCONNELL.C.R AND K.SOSIN (1984) "Some Determinants of Student Attitudes Towards Large Classes", *Journal of Economic Education* 15: 181-90.

[32] MORETTI.E (2004) "Workers' Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions", *American Economic Review* 94: 656-90.

[33] OYER.P (2008) "The Making of an Investment Banker: Macroeconomic shocks, Career Choice, and Lifetime Income", forthcoming *Journal of Finance.*

[34] PAGLIN.M AND A.RUFOLO (1990) "Heterogeneous Human Capital, Occupational Choice, and Male-Female Earnings Differences", *Journal of Labor Economics* 8: 123-44.

[35] PASCARELLA.E.T AND P.T.TERENZINI (1991) *How College Affects Students*, San Francisco: Josey-Bass.

[36] RIVKIN.S.G, E.A.HANUSHEK, AND J.F.KAIN (2005) "Teachers, Schools, and Academic Achievement", *Econometrica* 73: 417-59.

[37] ROCKOFF.J.E (2004) "The Impact of Individual Teachers on Student Achievement: Evidence From Panel Data", *American Economic Review* 94: 247-52.

[38] ROMER.D (1993) "Do Students Go To Class? Should They?", *Journal of Economic Perspectives* 7: 167-74.

[39] SACERDOTE.B (2001) "Peer Effects With Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics* 116: 681-704.

[40] STINEBRICKNER R. AND STINEBRICKNER T.R. (2008) "The Causal Effect of Studying on Academic Performance", *The B.E. Journal of Economic Analysis and Policy (Frontiers)* Volume 8, Issue 1, Article 14.

[41] WILLIAMS.D.D., P.F.COOK, B.QUINN AND R.P.JENSEN (1985) "University Class Size: Is Small Better?", *Research in Higher Education* 23: 307-17.

[42] ZIMMERMAN.D.J (2003) "Peer Effects in Academic Outcomes: Evidence From a Natural Experiment", *Review of Economics and Statistics* 85: 9-23.

# Table 1: Descriptive Statistics on Test Scores and Class Sizes

**Mean, between standard deviation in parentheses, within standard deviation in brackets**

| | Test Scores | | | Class Sizes | | |
|---|---|---|---|---|---|---|
| **Courses:** | **All** | **Core** | **Electives** | **All** | **Core** | **Electives** |
| **Mean** | 62.0 | 62.0 | 62.1 | 56.2 | 73.4 | 41.5 |
| **Standard Deviation Between Students** | (5.00) | (5.81) | (5.33) | (33.2) | (47.5) | (26.9) |
| **Standard Deviation Within Student** | [4.37] | [3.89] | [3.43] | [32.3] | [23.4] | [19.5] |

**Notes:** The decomposition of each statistic into the within and between variation takes account of the fact that the panel is unbalanced in each case. There are 40851 student-course level observations in total, covering 10873 students, and an average of 3.76 courses per student. Of these, 18923 (21928) observations are for core (elective) student-courses. The between and within standard deviations account for the panel being unbalanced.

## Table 2: Class Size Effects

**Dependent Variable: Test Score**

**Standard errors are clustered by course-year**

| | (1) Unconditional | (2) Controls | (3) Within Student | (4) Class Composition | (5) Faculty Dummies |
|---|---|---|---|---|---|
| **Class size** | -.011*** | -.011*** | -.012*** | -.013*** | -.015*** |
| | (.002) | (.002) | (.002) | (.002) | (.003) |
| **Implied Effect Size** | -.074*** | -.073*** | -.082*** | -.093*** | -.108*** |
| | (.016) | (.014) | (.014) | (.014) | (.020) |
| | [ -.106, -.042 ] | [ -.100, -.046 ] | [ -.107, -.056 ] | [ -.120, -.066 ] | [ -.148, -.069 ] |
| **Student fixed effect** | No | No | Yes | Yes | Yes |
| **Faculty dummies** | No | No | No | No | Yes |
| **Adjusted R-squared** | .006 | .016 | .574 | .575 | .618 |
| **Observations (clusters)** | 40851 (1775) | 40851 (1775) | 40851 (1775) | 40851 (1775) | 40851 (1775) |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. Observations are at the student-course-year level. Standard errors are clustered by course-year throughout. The dependent variable is the student's test score. In Columns 2 onwards we control for the number of faculty that teach on the course, the share of them that are full professors, the total number of credits obtained for completing the course, whether the course is a core or elective course for the student, and the share of the overall mark that is attributed to the final exam. In Column 4 we additionally control for the following course-year characteristics---the share of women, the mean age of students, the standard deviation in age of students, and the ethnic fragmentation among students, where the ethnicity of each student is classified as white, black, Asian, Chinese, and other, the fragmentation of students by department, the share of students who completed their undergraduate studies at the same institution, and the share of British students. In Column 5 we control for a complete series of faculty dummies, such that each faculty dummy is equal to one if faculty member j teaches on the course-year, and zero otherwise. The implied effect size is the effect of a one standard deviation increase in class size from its mean, on the test score, normalized by the standard deviation of test scores. The foot of each column reports the implied effect size, its standard error, and the associated 95% confidence interval. In Columns 1 and 2 these standard deviations are calculated over all students and classes, and in the remaining columns the standard deviations refer to the within student values.

## Table 3: Evaluations and Course Choice

| Dependent Variable: | Evaluations | | Choice | |
|---|---|---|---|---|
| | **(1) Postgraduate Course Satisfaction** | **(2) Undergraduate Course Satisfaction** | **(3) Choice of Core Courses** | **(4) Choice of Elective Courses** |
| **Class size** | -.002*** | .002 | -.024*** | -.016*** |
| | (.001) | (.001) | (.004) | (.004) |
| **Class size x number of available core courses on program** | | | .004*** | |
| | | | (.001) | |
| **Class size x number of available elective courses on program** | | | | .001 |
| | | | | (.000) |
| **Implied Effect Size at Mean** | | | -.090*** | -.071*** |
| | | | (.021) | (.013) |
| | | | [ -.131, -.050 ] | [ -.097, -.045 ] |
| **Implied Effect Size at 10th Percentile** | | | -.124*** | -.086*** |
| | | | (.022) | (.020) |
| | | | [ -.167, -.081 ] | [ -.126, -.046 ] |
| **Implied Effect Size at 90th Percentile** | | | -.053* | -.054*** |
| | | | (.027) | (.018) |
| | | | [ -.106, .000 ] | [ -.089, -.019 ] |
| **Fixed Effects** | Department | Department | Student | Student |
| **Adjusted R-squared** | - | - | .693 | .709 |
| **Observations (clusters)** | 104 | 90 | 18125 (710) | 21347 (1557) |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. In Columns 1 and 2 the dependent variable is the average student evaluation at the department-year level among postgraduate (undergraduate) students. This can range from 1 (least satisfied) to 4 (most satisfied). Observations are weighted by the number of student evaluation responses Panel corrected standard errors are calculated using a Prais-Winsten regression. This allows the error terms to be department specific heteroskedastic, and contemporaneously correlated across departments. A common AR(1) process is assumed for all departments. All observations are at the department-year level. Observations for 2 of the 23 departments in which students can be enrolled are dropped either because that department only offers its own courses in the last year of data, or because the department always offers all courses jointly with other departments. Hence the sample is based on a balanced panel of 21 departments over five academic years (1999/00-2003/4). The number of students enrolled in related departments is defined to be those students that are eligible to take any given course in the department for credit as part of their graduate degree program. We first calculate this enrolment for each module within the department, and then take its average over all courses within the department for each academic year. Class sizes are averages within the department-year. Hence in these columns we weight observations by the number of courses in the department that academic year. Weighted means of class size are then reported at the foot of the table. In Columns 3 and 4 the dependent variable is the student's test score on the course, observations are at the student-course year level, and standard errors are clustered by course-year. We control for the number of faculty that teach on the course, the share of them that are full professors, the total number of credits obtained for completing the course, the share of the overall mark that is attributed to the final exam, and the following course-year characteristics---the share of women, the mean age of students, the standard deviation in age of students, and the ethnic fragmentation among students, where the ethnicity of each student is classified as white, black, Asian, Chinese, and other, the fragmentation of students by department, the share of students who completed their undergraduate studies at the same institution, and the share of British students. The samples in Columns 3 (4) is restricted to those programs that have at least one required core (elective) course. The foot of each column reports the implied effect size, its standard error, and the associated 95% confidence interval. These standard deviations refer to the within student values. The effect size is also calculated at the 10th and 90th percentile of the number of available core (elective) courses on the programme that the student is enrolled in.

# Table 4: Non Linear Class Size Effects

**Dependent Variable: Test Score**

**Standard errors are clustered by course-year**

| | (1) Quintile | (2) Faculty Heterogeneity | (3a) Core | (3b) Electives |
|---|---|---|---|---|
| **Class size: quintile 2 [20-33]** | -.561*** | -.494*** | -1.16*** | -.339** |
| | (.128) | (.140) | (.336) | (.138) |
| **Class size: quintile 3 [34-55]** | -.971*** | -.949*** | -1.33*** | -.823*** |
| | (.147) | (.181) | (.368) | (.159) |
| **Class size: quintile 4 [56-103]** | -1.02*** | -1.08*** | -.968** | -.741*** |
| | (.188) | (.231) | (.461) | (.190) |
| **Class size: quintile 5 [104-211]** | -1.92*** | -1.97*** | -2.30*** | -1.54*** |
| | (.271) | (.338) | (.534) | (.331) |
| **Student fixed effect** | Yes | Yes | Yes | Yes |
| **Faculty dummies** | No | Yes | No | No |
| **Test: Quintile 2 = Quintile 3 (p-value)** | [.001] | [.004] | [.586] | [.001] |
| **Test: Quintile 3 = Quintile 4 (p-value)** | [.804] | [.513] | [.403] | [.669] |
| **Test: Quintile 4 = Quintile 5 (p-value)** | [.001] | [.003] | [.003] | [.016] |
| **Adjusted R-squared** | .575 | .618 | .697 | .713 |
| **Observations (clusters)** | 40851 (1775) | 40851 (1775) | 18923 (710) | 21913 (1557) |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. All observations are at the student-course year level. Standard errors are clustered by course-year. The dependent variable is the student's test score. In all columns we control for the number of faculty that teach on the course, the share of them that are full professors, the total number of credits obtained for completing the course, whether the course is a core or elective course for the student, and the share of the overall mark that is attributed to the final exam. In all Columns we control for the following course-year characteristics---the share of women, the mean age of students, the standard deviation in age of students, and the ethnic fragmentation among students, where the ethnicity of each student is classified as white, black, Asian, Chinese, and other, the fragmentation of students by department, the share of students who completed their undergraduate studies at the same institution, and the share of British students. In Column 2 we additionally control for a complete series of faculty dummies, such that each faculty dummy is equal to one if faculty member j teaches on the course-year, and zero otherwise. Column 3a (3b) restricts the sample to core (elective) courses only.

## Table 5: Heterogeneous Non Linear Class Size Effects

**Dependent Variable: Test Score**

**Standard errors are clustered by course-year**

| | (1) British | (2) Undergraduate Institution | (3) House Prices |
|---|---|---|---|
| Group 0 = | Non British | Different | Below median |
| Group 1 = | British | Same | Above median |
| Class size: quintile 2 [20-33] \| group = 0 | -.554*** | -.479*** | -.459*** |
| | (.147) | (.146) | (.158) |
| Class size: quintile 2 [20-33] \| group = 1 | -.236 | -.568*** | -.317 |
| | (.216) | (.213) | (.233) |
| Class size: quintile 3 [34-55] \| group = 0 | -.954*** | -.911*** | -.916*** |
| | (.185) | (.185) | (.202) |
| Class size: quintile 3 [34-55] \| group = 1 | -.974*** | -1.16*** | -1.08*** |
| | (.259) | (.250) | (.265) |
| Class size: quintile 4 [56-103] \| group = 0 | -1.12*** | -1.04*** | -1.13*** |
| | (.238) | (.233) | (.249) |
| Class size: quintile 4 [56-103] \| group = 1 | -.891*** | -1.29*** | -.940*** |
| | (.282) | (.315) | (.305) |
| Class size: quintile 5 [104-211] \| group = 0 | -1.92*** | -1.95*** | -2.31*** |
| | (.340) | (.340) | (.370) |
| Class size: quintile 5 [104-211] \| group = 1 | -2.36*** | -2.12*** | -1.98*** |
| | (.432) | (.414) | (.414) |
| P-value: Equal class size effects: quintile 2 | [.130] | [.657] | [.520] |
| P-value: Equal class size effects: quintile 3 | [.926] | [.248] | [.465] |
| P-value: Equal class size effects: quintile 4 | [.336] | [.297] | [.416] |
| P-value: Equal class size effects: quintile 5 | [.137] | [.526] | [.269] |
| Student fixed effect and faculty dummies | Yes | Yes | Yes |
| Adjusted R-squared | .618 | .618 | .628 |
| Observations (clusters) | 40851 (1775) | 40851 (1775) | 27238 (1735) |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. Observations are at the student-course-year level. Standard errors are clustered by course-year throughout. The dependent variable is the student's test score. Throughout we control for the number of faculty that teach on the course, the share of them that are full professors, the total number of credits obtained for completing the course, and the share of the overall mark that is attributed to the final exam, and a complete series of faculty dummies, such that each faculty dummy is equal to one if faculty member j teaches on the course-year, and zero otherwise. In all columns we also control for the following course-year characteristics---the share of women, the mean age of students, the standard deviation in age of students, and the ethnic fragmentation among students, where the ethnicity of each student is classified as white, black, Asian, Chinese, and other, the fragmentation of students by department, the share of students who completed their undergraduate studies at the same institution, and the share of British students. The sample in Column 3 is restricted to students who do not reside in university accommodation. We classify a student as living in a three digit postcode sector with above median house prices if the average sale price of flats/marionettes in the postcode in the year of study is above the median sale price among all three digit postcodes that year. The implied effect size is the effect of a one standard deviation increase in class size from its mean, on the test score, normalized by the standard deviation of test scores. The foot of each column reports the p-value on the null hypothesis that the coefficient on each quintile of class size is the same in groups 0 and 1.

# Table A1: Departmental Inputs and Student Enrolment

**Prais-Winsten Regression Estimates**

| Dependent Variable: | Numbers of Teaching Faculty | | | | Class Size | |
|---|---|---|---|---|---|---|
| | **(1) All Faculty** | **(2) Full Professors** | **(3) Other Professors** | **(4) Non Professors** | **(5a) Mean** | **(5b) Mean** |
| **Number of students enrolled in department** | .057** | .026** | .038* | -.007 | .112*** | .110*** |
| | (.029) | (.010) | (.021) | (.007) | (.014) | (.016) |
| **Number of students enrolled in related departments** | -.006 | -.000 | -.004 | -.000 | .025** | .026** |
| | (.012) | (.005) | (.007) | (.002) | (.011) | (.011) |
| **Number of courses offered by department** | | | | | | .044 |
| | | | | | | (.030) |
| **Number of programs offered by department** | | | | | | -.011 |
| | | | | | | (.418) |
| **Mean of dependent variable** | 13.4 | 4.27 | 7.56 | 1.43 | 26.4 | 26.4 |
| **Test: equal enrolment effects (p-value)** | [.024] | [.030] | [.050] | [.410] | [.000] | [.000] |
| **Number of own department enrollees required to increase outcome by one unit** | 17.7 | 37.9 | 26.4 | -145 | 8.95 | 9.09 |
| **Number of related department enrollees required to increase outcome by one unit** | - | - | - | - | 40.0 | 37.8 |
| **Department fixed effects** | Yes | Yes | Yes | Yes | Yes | Yes |
| **Observations (department-year)** | 105 | 105 | 105 | 105 | 105 | 105 |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. Panel corrected standard errors are calculated using a Prais-Winsten regression. This allows the error terms to be department specific heteroskedastic, and contemporaneously correlated across departments. A common AR(1) process is assumed for all departments. All observations are at the department-year level. Observations for 2 of the 23 departments in which students can be enrolled are dropped either because that department only offers its own courses in the last year of data, or because the department always offers all courses jointly with other departments. Hence the sample is based on a balanced panel of 21 departments over five academic years (1999/00-2003/4). The number of students enrolled in related departments is defined to be those students that are eligible to take any given course in the department for credit as part of their graduate degree program. We first calculate this enrolment for each module within the department, and then take its average over all courses within the department for each academic year. All faculty numbers refer to faculty that teach on at least one course during the academic year. In Column 4 non professors refers to teaching faculty that do not have a doctorate degree. In Columns 5a and 5b, class sizes are averages within the department-year. Hence in these columns we weight observations by the number of courses in the department that academic year. Weighted means of class size are then reported at the foot of the table. At the foot of the table we also report the p-value of a t-test of equality of the coefficients on own department and outside department enrolments. We also report the implied inverse of the coefficients on own department and outside department enrolments to calculate the change in these variables that are associated with a one unit increase in each dependent variable.

## Table A2: Robustness Checks

**Dependent Variable: Test Score**

**Standard errors are clustered Columns 1 and 3, and robustly estimated in Column 2**

| | Course Omitted Variables | | |
|---|---|---|---|
| | **(1) Course Difficulty** | **(2) Course Fixed Effects** | **(3) Time Variation** |
| **Class size** | -.015*** | -.006* | -.014** |
| | (.003) | (.004) | (.007) |
| **Share of students that are re-sitting the course** | .747 | | |
| | (2.75) | | |
| **Implied Effect Size** | -.108*** | -.042* | -.045** |
| | (.020) | (.025) | (.023) |
| | [ -.148, -.069 ] | [ -.091, .007 ] | [ -.090, -.001 ] |
| **Fixed Effect** | Student, Faculty | Course | Course |
| **Adjusted R-squared** | .618 | .089 | .630 |
| **Observations (clusters)** | 40851 (1775) | 40851 | 1775 (116) |

**Notes:** *** denotes significance at 1%, ** at 5%, and * at 10%. In Columns 1 and 2 the dependent variable is the student's test score on the course, and observations are at the student-course year level. In Column 3 the dependent variable is the average test score across students on the course, observations are at the course-year level. In Column 1 (3) the standard errors are clustered by course-year (department-year) and in Column 3 they are robustly estimated. We control for the number of faculty that teach on the course, the share of them that are full professors, the total number of credits obtained for completing the course, the share of the overall mark that is attributed to the final exam, and the following course-year characteristics---the share of women, the mean age of students, the standard deviation in age of students, and the ethnic fragmentation among students, where the ethnicity of each student is classified as white, black, Asian, Chinese, and other, the fragmentation of students by department, the share of students who completed their undergraduate studies at the same institution, and the share of British students. In Column 1  we additionally control for a complete series of faculty dummies, such that each faculty dummy is equal to one if faculty member j teaches on the course-year, and zero otherwise. In Columns 2 and 3 we control for a complete series of course fixed effects rather than student fixed effects. The foot of each column reports the implied effect size, its standard error, and the associated 95% confidence interval.
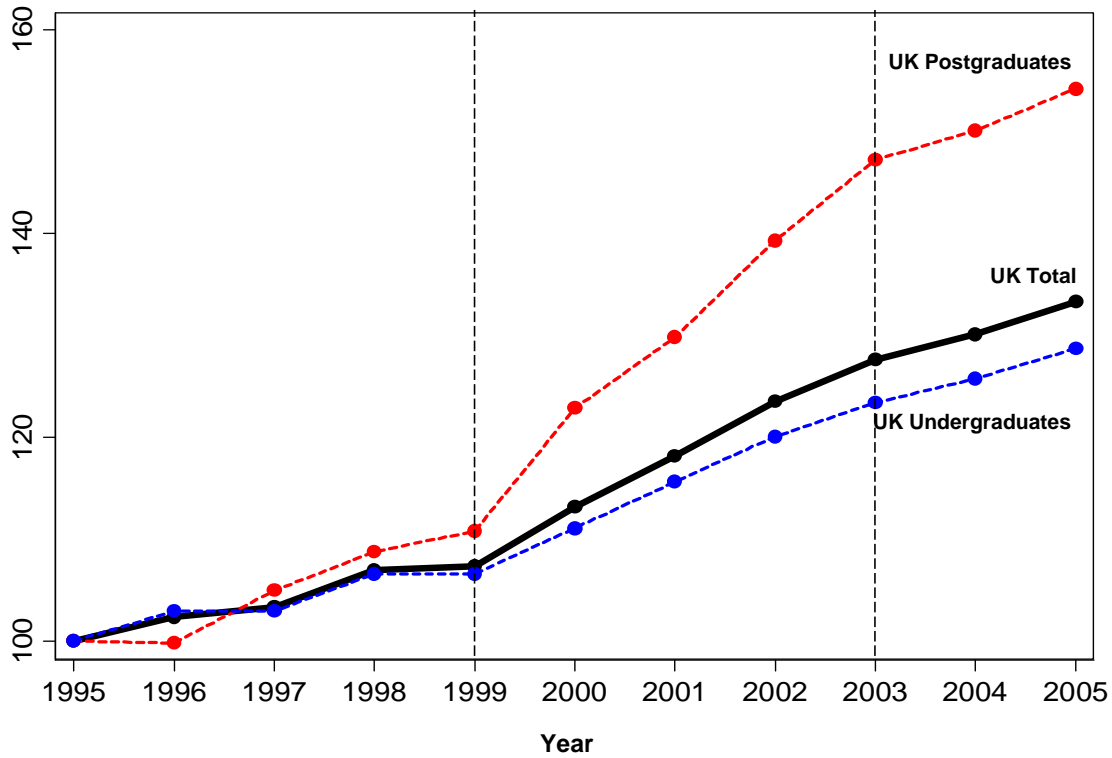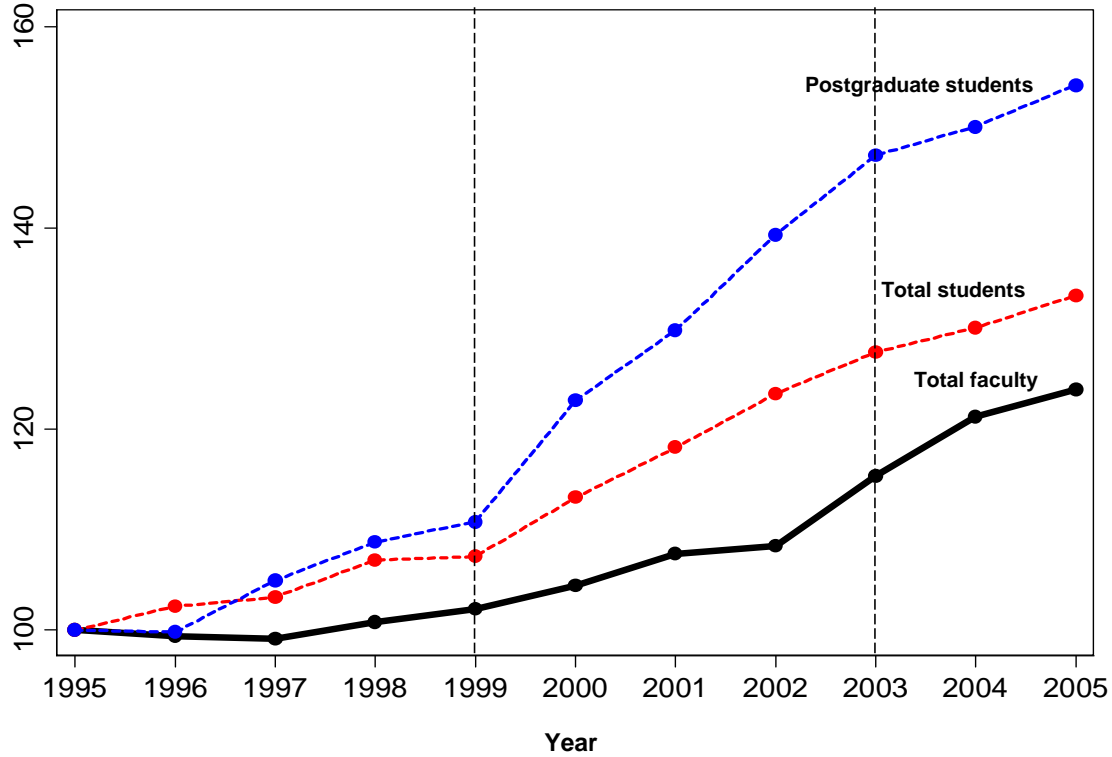
## Fig. 1A: Aggregate UK Enrolment in Higher Education



## Fig. 1B: Aggregate Numbers of Faculty in Higher Education



**Notes:** Figure 1A shows total full time equivalent student enrolments, in the UK, and split for the UK by undergraduate and postgraduate students. Postgraduate students correspond to those in their fourth or higher year of tertiary education. Figure 1B shows total full time equivalent faculty numbers in higher education in the UK, and the total full time equivalent student enrolment, and that for postgraduate students. Postgraduate students correspond to those in their fourth or higher year of tertiary education. Each time series is normalized to be equal to 100 in 1995. The data source for these figures is the Higher Education Statistics Agency (http://www.hesa.ac.uk/), accessed on 15 November 2007.

## Fig. 2: Distribution of Classmarks, by Course-Year



**Course-Year: Sorted by Ascending Proportion of B's**
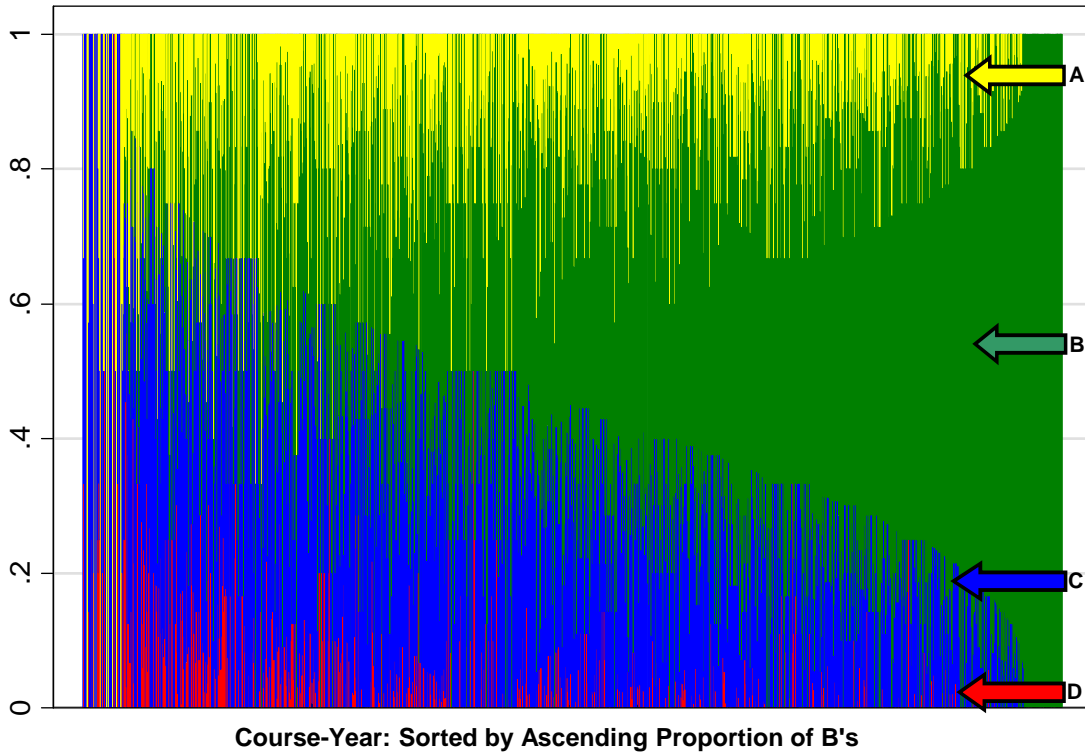
## Fig. 3A: Within Student Variation in Test Score



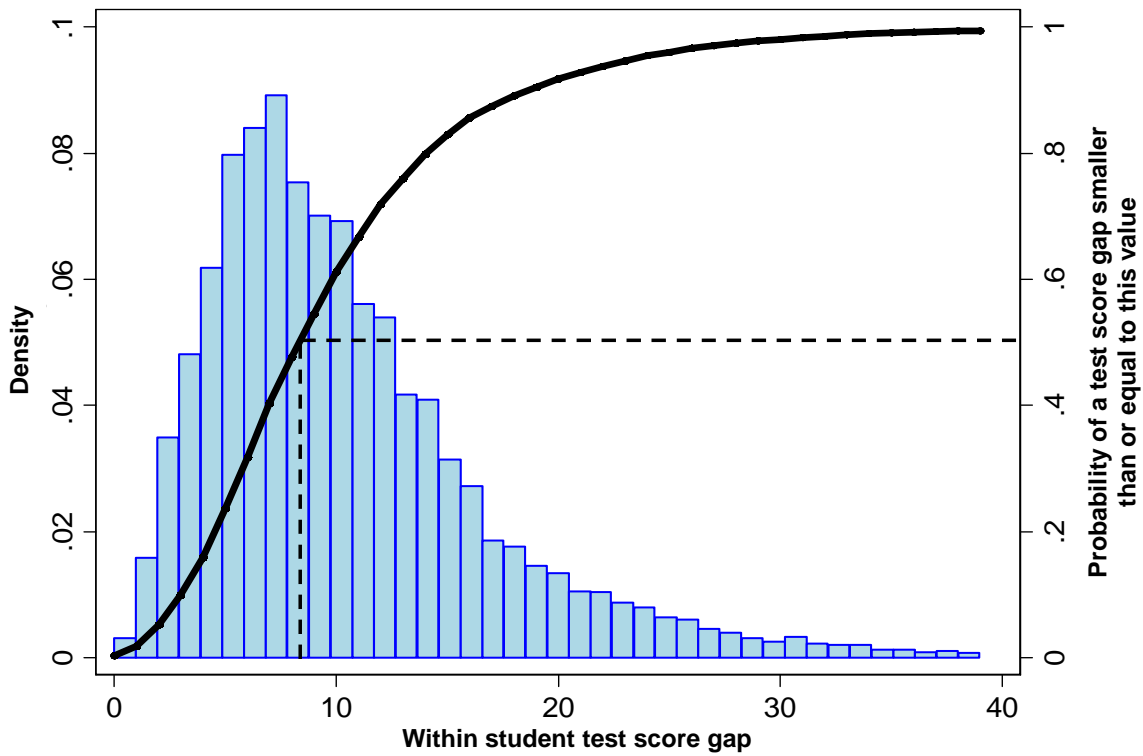**Within student test score gap**

**Notes:** In Figure 2, each vertical line represents one course-year. The figure shows for each course-year, the proportions of students that obtain each classmark. These classifications correspond to final exam marks greater than or equal to 70 for an A-grade, between 60 and 69 for a B-grade, between 50 and 59 for a C-grade, and below 50 for a D-grade (fail). To ease exposition, the course-years are sorted into ascending order of the share of B-grades awarded. In Figure 3A, the within student test score gap is defined to be the difference in the student's maximum and minimum test scores across courses.

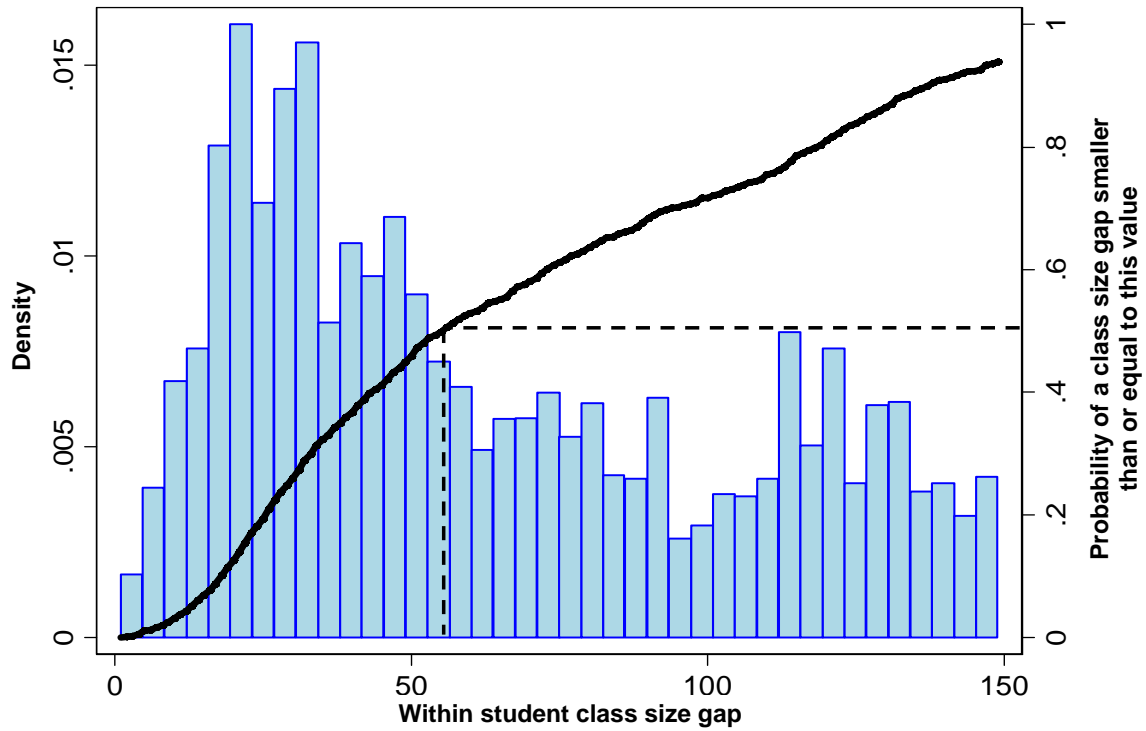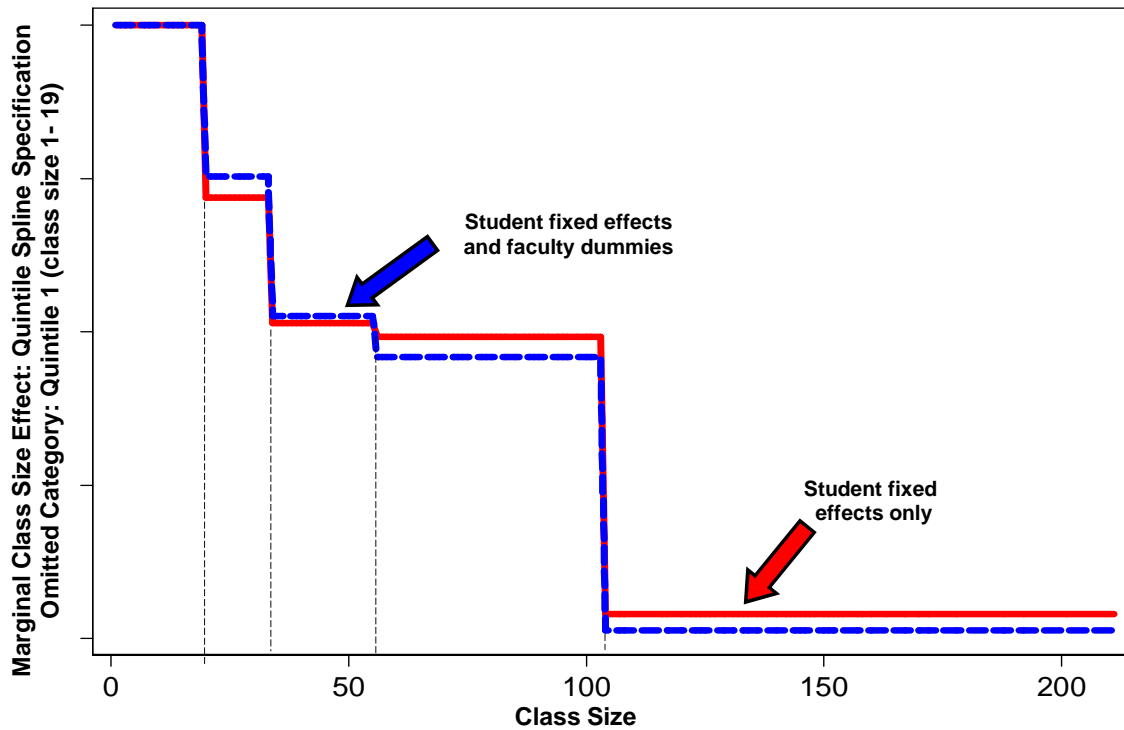## Fig. 3B: Within Student Variation in Class Size



## Fig. 4: Non-Linear Class Size Effects



**Notes:** In Figure 3B, the within student class size gap is analogously defined. The right hand axis on each figure shows the probability that a given student has a test score (class size) gap less than or equal to the value on the horizontal axis. On each figure, we indicate the test score and class size gaps for the median student in the sample. Figure 4 plots the coefficients from a panel data spline regression of test scores on a series of dummies for whether the class size is in any given quantile of the class size distribution, and the same controls as described in the Tables. The two sets of estimates correspond to the specifications with and without faculty dummies. The omitted category is class sizes in the first quintile, corresponding to class sizes of 1 to 19. The second quintile corresponds to class sizes from 20 to 33, the third from 34 to 55, the fourth from 56 to 103, and the fifth to 104 to 211.

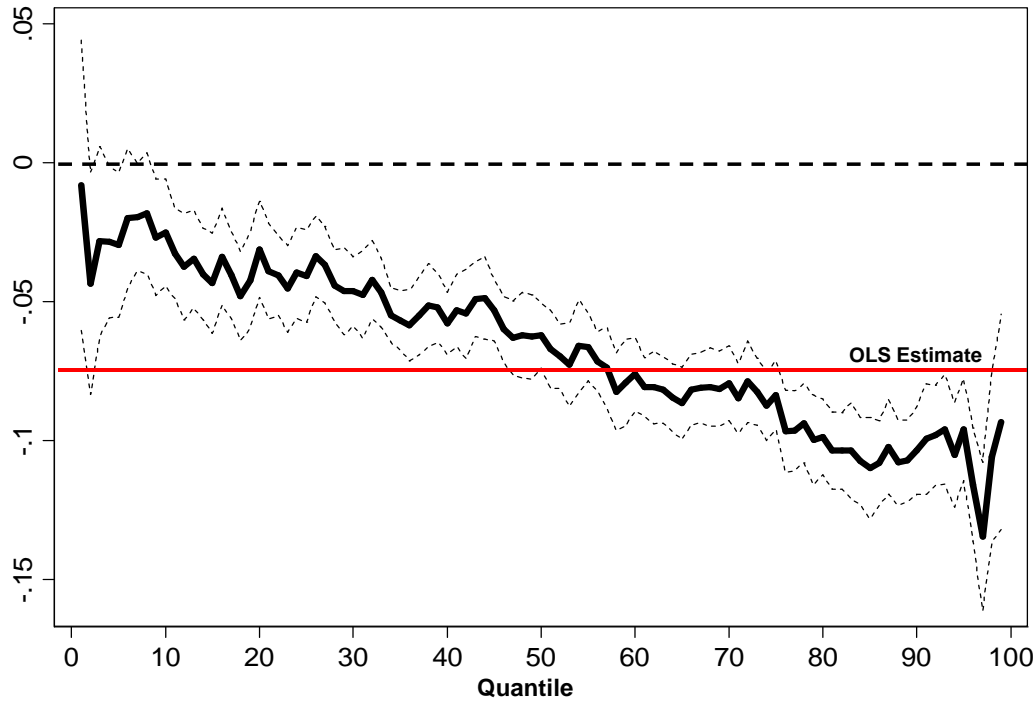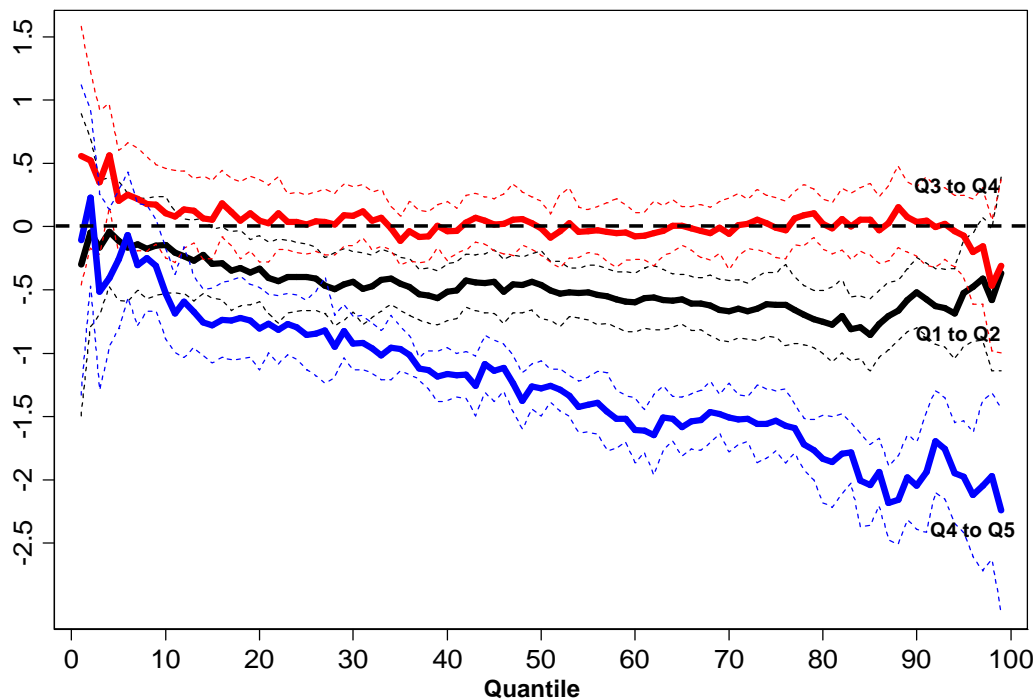## Fig. 5A: Quantile Regression Estimates of the Effect Size



## Fig. 5B: Quantile Regression Estimates of Moving from One Quintile to the Next in the Class Size Distribution



**Notes:** Figure 5A graphs the estimated effect size on the test score at each quantile of the conditional distribution of student test scores, and the associated 95% confidence interval. Figure 5B graphs the estimated effect of moving from quintile q to quintile q' in the distribution of class sizes on the test score at each quantile of the conditional distribution of student test scores, and the associated 95% confidence interval. In each case, the distribution of test scores is conditioned on the number of faculty that teach on the course, the share of them that are full professors, the total number of credits obtained for completing the course, and the share of the overall mark that is attributed to the final exam, and the following course-year characteristics---the share of women, the mean age of students, the standard deviation in age of students, and the ethnic fragmentation among students, where the ethnicity of each student is classified as white, black, Asian, Chinese, and other, the fragmentation of students by department, the share of students who completed their undergraduate studies at the same institution, and the share of British students. For expositional ease we do not show the estimates of moving from quintile 2 to quintile 3. In line with the results in Table 4, these are slightly more negative than moving from quintile 1 to quintile 2 at each quantile.