Centre for Market and Public Organisation
Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol BS8 1TX
http://www.bris.ac.uk/Depts/CMPO/

Tel: (0117) 33 10799
Fax: (0117) 33 10705
E-mail: cmpo-office@bristol.ac.uk

The Impact of Classroom Peer Groups on Pupil GCSE Results

Adele Atkinson, Simon Burgess, Paul Gregg, Carol Propper and Steven Proud

January 2008

Working Paper No. 08/187

ISSN 1473-625X

CMPO Working Paper Series No. 08/187

The Impact of Classroom Peer Groups on Pupil GCSE Results

Adele Atkinson, Simon Burgess, Paul Gregg, Carol Propper
and
Steven Proud

CMPO, University of Bristol

January 2008

Abstract

The effect of a more able peer group on a child's attainment is considered an integral part in estimating a pupil level educational production function. Examinations in England at age 16 are tiered according to ability, leading to a large stratification of pupils by ability. However, within tiers, there is a range of policies between schools regarding setting, ranging from credibly random to strict setting by results from examinations at age 14. We use this variation to estimate ordinary least squares (OLS) estimates, with school and teacher fixed effects, of the effect of a more able peer group using a subset of schools that has apparently random allocation of pupils. As a robustness test of the apparently random setting results, we use an instrumental variables (IV) methodology developed by Lefgren (2004b). We find significant, positive, and non-trivial effects of a more able peer group using both the OLS and IV estimations for English and mathematics.

Keywords: peer groups, education

JEL Classification: J13, D1, I21, I38

Electronic version: http://www.bris.ac.uk/Depts/CMPO/workingpapers/wp187.pdf

Acknowledgements

We would like to thank the schools who participated in the CMPO school survey and allowing access to linked teacher and class data. Also we would like to thank the Leverhulme Trust and the ESRC for financial support on this project. Thanks also to Deborah Wilson and seminar participants at the University of Bristol for comments and suggestions.

Address for Correspondence
CMPO, Bristol Institute of Public Affairs
University of Bristol
2 Priory Road
Bristol
BS8 1TX
Steven.Proud@bristol.ac.uk
www.bris.ac.uk/Depts/CMPO/

CMPO is jointly funded by the Leverhulme Trust and the ESRC

1. Introduction

There has long been an interest in the effect of a pupil's peers upon their outcomes. Britain, like all nations, has geographic areas of relative deprivation and affluence. Access to schools by catchment areas (residential location), academic selection or parental choice mechanisms all result in large variations in the pupil mix within schools (see Burgess et al. 2004). Whether the sorting of children in these ways have an impact on a child's outcomes is thus a key and longstanding policy concern. If there is a significant effect of a more able peer group, then stratification of pupils into ability based teaching groups may lead to a polarisation of the population, with more able students only helping the similarly able. However, there is a standard problem when comparing pupil attainment across schools according to the school mix. Schools with intakes with low measured ability on average are likely to be attracting pupils that have unmeasured adverse characteristics influencing their future achievement prospects. These pupils may achieve less in the future, even given their initial measured ability, for reasons relating to their home or school characteristics rather than the mix of pupils within the classroom. The correlation of the unmeasured attributes with both the outcome measure and the peer group indicator results in an omitted variable bias that likely overstates the influence of the peer group.

Isolating the influence of the peer group from unobserved heterogeneity is not straightforward, but there has been a rapid growth in studies attempting this by econometric techniques and experimental policy design, such as Lefgren (2004b) and others. In the UK there have been no true experimental studies capable of addressing this issue. However, there have been a small but growing number of studies addressing this issue using other techniques, such as Gibbons and Telhaj (2006)

We use data from key stage 3 (KS3), examined at age 14 and GCSE, examined at age 16[1]. The GCSE qualification has two or three levels of difficulty of examination, or tiers, that pupils can be entered for within each subject. This encourages schools to group students into sets by ability for these examinations. Within each individual examination tier, however, there is much greater variation in setting ranging from strict

[1] The structure of the English school system is discussed in more detail in the data section.

ordering to apparent random allocation. We use a unique dataset for England[2], containing data on the sets in which pupils are taught at ages 15 and 16, giving us data on the peer group that the students directly experience. We estimate the effect of a more able peer group using a sample of schools with a credibly random distribution of pupils, and utilise an instrumental variables technique developed in Lefgren (2004b), and using these methods, we find significant and non trivial positive effects of a more able peer group, which are smaller than the ordinary least squares estimates. Section 2 summarises the recent literature. Section 3 discusses identification issues. Section 4 examines the data. Section 5 discusses the results. Section 6 offers some concluding discussion.

2. Literature

The literature looking at peer group effects is long standing. Perhaps the most important was the Coleman Report (Coleman et al (1966)), which argues that the "Attributes of other students account for far more variation in the achievement of minority group children than do any attributes of school facilities and slightly more than do attributes of staff" Coleman et al (1966, page 86). More recent literature in this vein shows substantial correlations between a child's outcomes and that of their peers. Jencks and Mayer (1990) point out that pupils are more likely to drop out of school if their peers are of lower socioeconomic status. Mayer (1991) demonstrates that pupils at a school with peers having a higher socioeconomic status are less likely to drop out of school between the tenth and twelfth grades, whilst white pupils attending schools with mainly black or Hispanic peers are also more likely to drop out of school early. However, this literature makes little attempt to isolate the peer group effect from school or teacher effectiveness or from biases from unobserved pupil attributes.

Some caution therefore needs to be applied when drawing conclusions from these studies. It is probable that the peer measure is correlated with the error term in the regression. This correlation could come from the fact that selection into schools and allocation of teachers to classes and the allocation of pupils to classes are most definitely not random, and are in fact assigned based upon ability and other (unobserved) characteristics. Similarly, as discussed in Manski (1993) these effects

[2] Data collected to examine the effects of the introduction of performance related pay. See Atkinson et al (2004)

may simply be due to pupils experiencing the same circumstances that could similarly affect all of their outcomes. This could be in the form of a particularly good (or bad) classroom teacher, or some other common shock. These shocks are often unobservable, and so cannot be directly controlled for. Thus, we would expect the ability of a pupil in a classroom environment to be implicitly correlated with that of their peers. Within neighbourhoods, we may expect to see people living with similar characteristics. People may move into an area because of the people who live there, and similarly, richer parents may be able to buy housing in areas with better schools and better ability pupils.

There is now a small but rapidly growing body of literature that uses a variety of solutions to address this problem, but to date this literature is mainly US based. A number of studies look at randomly allocated accommodation within higher education in the US. Sacerdote (2001) uses the fact that students are randomly assigned a dorm-mate upon arrival at Dartmouth College and find that peers have an effect on grade point average. This same argument is used by Zimmerman (2003) and Winston and Zimmerman (2004), both of which find no credible effect on the top of the SAT distribution, but instead find evidence of a negative effect on mid-ability students grouped with students in the bottom 15% of the SAT distribution. Stinebrickner and Stinebrickner (2006) use this method with data from first year students at Berea College, and find significant effects on female students' results from their roommate's high school grade point average and their family income.

More relevant for our purposes are studies looking at school age children. Angrist and Lang (2004) use results from the METCO program in Boston, which sends mostly black students out of Boston's public schools into the more affluent suburbs. They assess the impact of the low reading score of the transferred students on the students in the receiving districts. They suggest that the differences in peer groups this generates show small negative localised effects on ethnic minority pupils in the host districts. Hoxby (2000) exploits the variation in gender and racial mix within schools across time suggesting that "a credible [positive] exogenous change in [school] peers' reading scores raises a students own score between 0.15 and 0.4 points" (Hoxby (2000), 2). The effects she identifies are thus quite substantial. Proud (2008) uses a similar strategy to Hoxby, and shows a change in the gender make up of the peer group does not make a good proxy for ability, and instead considers the proportion of girls to affect other aspects such as behaviour and teacher-pupil interactions. Lavy and Schlosser (2007) demonstrate that a higher number of female peers leads to higher academic outcomes, but they argue that this is not due to an increase in the ability of the peer group, but rather due to behaviour within the classroom due to the compositional make-up of the school.

Yet these studies are based on school intake variations, whereas the majority of peer effects are more likely to be motivated in the classroom. The probable non-random assignment of pupils and/or teachers to classes, or common patterns of attainment development within a class other than from the peer group means that biases may remain. To address these issues Hanushek et al (2003) use lagged student attainment to net out individual fixed effects which, along with school and school-grade fixed effects, could account for the systematic but unobserved differences in students and schools. They also deal with any common production of added value in the school grade (other than peer effects) by using lagged peer attainment. Their results suggest that, having controlled for school characteristics in a value added model, there are no mean peer attainment effects but there are adverse effects from having more poor children in the class or less homogeneity of ability within a class. However, introducing student fixed effects, school fixed effects and school-grade fixed effects produces a consistent pattern, that a greater proportion of poorer children is not detrimental and actually weakly positive. Betts and Zau (2004) also use student fixed effects are draw similar conclusions. There remain concerns though that non-random allocation of pupils to classes within schools will also be correlated with non-random allocation of teaching resources including teacher quality to classes. Thus, there is the danger that teacher ability is correlated with mean prior attainment of peers in the class.

Burke and Sass (2006) go one step further introducing teacher fixed effects, in addition to school fixed effects, into value added models looking at peer effects in Florida middle schools. The broad pattern of results is that positive peer effects disappear once a value added model is adopted, with school and teacher fixed effects making little further difference. Lefgren (2004b) uses an instrumental variables estimation technique in order to remove the correlation with the error term. As an instrument he uses the variation in class setting policies in Chicago secondary schools. He argues this decision is school level and unrelated to pupil characteristics after inclusion of school fixed effects, but does alter the pupil composition of classes. This is measured by the R-squared obtained by regressing the pupils' results against a set of class dummies. This results in smaller, but still significant, effects than non-IV estimates. Clark (2003) also uses an instrumental variable estimation technique looking for the effect of being placed with older pupils in a middle school environment rather than a primary school followed by a secondary school. He uses a binary dummy based on whether they attend a middle school as an instrument for peer behaviour and finds significant peer effects on behaviour.

Robertson and Symons (2003) use UK data to estimate a production function using peer group, parental input and schooling as its inputs. They examine the effect of streaming on children's outcomes, and show that children placed in the top stream within a school benefit, whilst children placed in the bottom stream suffer.

Further studies look at variations in peer group effects across population sub-sets. On effects of different levels of ability, Summers and Wolfe (1977) find that low achieving pupils benefited from being in a school with high achievers whereas the high ability pupils are not significantly affected. This positive effect to the lower ability pupils is backed up by Dills (2005) who consider the introduction of a magnet school into a school district that cream off the best students from the schools, and found that the lower ability pupils' performances are lowered following the removal of the high ability pupils. Similarly, Henderson et al (1978) show that a mixing of weak and strong students will benefit the overall student population, but at the cost of lowering the outcomes for the higher achievers. Similar results are also found by Bradley and Taylor (2007), who use pupils moving between schools to address the problems inherent with estimating peer effects, and find the effects of a more able peer group are stronger for low ability students than for higher ability students. However, Betts and Shkolnik (2000) find little evidence of differential effects of ability grouping for high or low ability pupils. Further to this, Gibbons and Telhaj (2006) examine pupil attainment at age 14 in England, and find that whilst middle and high ability pupils have the same response to an improvement in the ability of the peer group, the lowest ability pupils do not seem to gain much, if anything.

In summary, the literature shows, on the whole, that being grouped with a higher ability peer group leads to an increase in outcomes. However, the effects found in early studies are likely to have mis-estimated the effects. In this study, we consider the effects, within classrooms in England, of a more able peer group. We use a measure of grouping within the school, within the tier of examination entered for at GCSE to estimate these effects for schools with a credibly random grouping strategy. Furthermore, we validate these results using the methodology developed by Lefgren (2004b).

3. Identification of peer effects

Researchers face substantial problems with how to correctly identify the peer effects. This may be due to the non-random allocation of pupils to schools, and within the schools into classes and classes to teachers. A pupil's peers within a school are often likely to have a similar social background due to fixed catchment areas for schools. That is, housing in the catchment areas of good schools is likely to be more expensive due to higher demand and is thus available to richer parents (Gibbons and Machin, 2003). Similarly, fee paying, religious and selective schools are also likely to have pupils with similar demographics, whilst also potentially having better facilities and resources available to them. So it is important to control for these school entry effects by including school-year or pupil fixed effects. Within the school there is also likely to be a non-random assignment of peers within classes. This is especially true for GCSE classes in the UK as pupils are often assigned to classes based on previous exam results as well as potentially ability that is unobservable to researchers based on teacher's assessment

A natural place to start in considering pupil attainment at GCSE level (age 16) is the general cumulative education production function developed by Todd and Wolpin (2003):

$$GCSE_{ij} = A_t[F_{ij}, S_{ij}, \mu_j, \varepsilon_{ij}]$$

where GCSE is the exam result for the pupil (*i)* in each subject (*j*) considered, A is the cumulated achievement function with F and S representing the entire input histories of the family and schools over the child's life to date, as they apply to subject j; μ is a composite variable representing individual time invariant characteristics such as ability to learn the subject and ε captures any measurement error.

On the assumption that past inputs and the past attainment stemming from the individual endowment can cumulated into a lagged attainment measure this can be re-written as:

$$GCSE_{ij} = \beta_1 F_{ijt} + \beta_2 S_{ijt} + \lambda A_{jt-2} + v_{ij} + \eta_{jit}$$

To explore class based peer group effects we can look to split up the school inputs component into:

$$S_{ijt} = C_{ijt} + T_{jt} + S_t$$

where C is the measure of the class peer group for each individual in each subject, T represents teacher quality inputs for each subject and S is the residual school level inputs reflecting school ethos, administration etc. which does not vary across individuals or classes. To avoid contamination of the peer group with any other co-produced attainment at class level during the two year GCSE course, the peer group measure is measured as the average outcome of the peer group 2 years previously. So here, our measure of the peer group (*classave*) is the mean Key Stage 3 (KS3) score of the set, not including the subject child, where KS3 is taken at the end of school year preceding the GCSE courses starting. In our estimation vectors of school-year fixed effects are included to capture school level variations in school effectiveness and we also explore teachers fixed effects. More detail on this in given in the data section.

Even with well measured pupil prior attainment, school, and teacher fixed effects there will remain a concern that a measure of class level peer group effects may still be biased. The data reveals, discussed in detail later, that the extent of grouping into sets by ability at the class level varies from school to school, especially in English. Within mathematics, on the other hand, pupils are widely grouped into sets with peers of very similar ability. In order to get a measure of the extent of setting at the school level, we regress the observed ability measure, that is the key stage 3 score, against a set of class dummies, and obtain the R-squared value. This gives us a measure to which a pupil is grouped in a class with pupils of a similar ability to themselves. We try to use this R-squared value in order to consider schools that use a credibly random policy for assigning their students to teaching sets. That is, if a school has an R-squared close to 1, then a pupil will be taught in a class with pupils with very similar scores at key stage 3, whilst if a school has an R-squared close to zero, then there will be more random assignment of pupils according to their key stage 3 score.

However, at GCSE, pupils are entered for a tier of examination, depending on how the school expects them to perform. Within these tiers, only a sub-sample of the grades is available to the students. This is likely to generate setting policy as the teacher may well find it easier to teach to a specific standard script rather than have students entering different exam tiers in the same class. So we consider regressions dependent on the tier the pupils are entered for at GCSE. Within tier, we also create an R-squared measure of setting within tier within the school. Since the curriculum taught in British secondary schools is regulated by the national curriculum, we hope that pupils within each tier will be taught in similar ways to each other, thus removing one of the potential alternative mechanisms for the peer effects to operate. We show in the data section there is much more evidence of apparently random setting practices within tier than exists within a whole school. However, in these regressions by tier entry, we only consider those schools that have 2 or more sets entered for any particular tier in order to compare the results within the school. So, whilst we may not find much evidence of credibly random distribution of pupils according to ability within schools, we do find evidence of credibly random distribution of pupils within the tier that the pupils are entered for. Due to this tiering policy, it is unlikely that many, if any, schools have a credibly random setting policy for the whole school, but our identification strategy should allow us to use schools with credibly random strategies within the examination tier they are entered for.

In order to estimate the effect of a more able peer group within the educational production function described above, we estimate

$$GCSE_{ij_t} = \alpha + X_{ij}\delta_j + classave_{ij_{t-2}}\gamma_j + s_t + t_t + \varepsilon_{ij} \qquad (1)$$

where GCSE is the GCSE score for pupil *i* at time *t* in subject *j*, *classave* is the mean of the peer group's key stage 3 score, not including child *i,* whilst *X* includes exogenous pupil level demographics. We include school–year fixed effects that have the effect of removing any other effects that may be constant across the pupil data within the cohort entry to the school. Further we explore the implications of adding teacher fixed effects.

We would like to be able to remove as much unobserved heterogeneity from the model as possible by including as much relevant background for the children as is available. However, in our dataset, there is little data regarding the child's background and other details. We try to reduce this unobserved heterogeneity by also including other measures of the pupil's ability in the form of the pupil's key stage 3 scores from other subjects, as well as demographics including their age within year, gender and a measure of the deprivation of their postcode from the index of multiple deprivation, from the Department of the Environment, Trade and Regions of local area deprivation.

In order to identify the effect of the peer ability score, we need to consider schools that use a credibly random setting policy. In order to do this, we consider schools that have an R-squared score of less than 0.35[3] within the tier as having a credibly random distribution of pupils. Schools that have an R-squared score of greater than 0.4 are defined as having a large R-squared, and do not have a credibly random distribution of pupils. This cut off is essentially arbitrary but the sample sizes within tier start to get very small, especially for maths but we can check for any residual bias by comparing the results those for the IV approach of Lefgren (2004b).

Whilst we may be confident that these schools have a random distribution of pupils to classes, in order to validate this, we appeal to the methodology developed in Lefgren (2004b**)** This identification strategy utilises the same R-squared measure as defined above, interacted with the pupil's subject specific key stage 3 score to estimate the following two-stage regression for each subject examined at key stage 3; English and maths:

$$classave_{ij_{t-2}} = a + X_{ij}\beta_j + R^2_{sj}KS3_{ij_{t-2}}\psi_j + s_t + t_t + u_{ij} \qquad (2)$$

$$GCSE_{ij_t} = \alpha + X_{ij}\delta_j + classave_{ij_{t-2}}\gamma_j + s_t + t_t + \varepsilon_{ij} \qquad (3)$$

[3] Whilst this may be considered a large cut-off, we are limited by having a small sample of schools, and so we must allow for schools with a less strict definition of random distribution. However, we appeal to the methodology of Lefgren (2004b) to verify these results.

where R^2 is the setting measure and KS3 is the pupil's own key stage 3 measure[4].

This IV approach, in line with Lefgren (2004b) only allows the use of one measure of peer ability, and we adopt the most common representation, that is the average of a pupil's classroom peer's lagged attainment score at 14 (Key Stage 3 scores). There are many potential mechanisms for peer effects to operate. A pupil may benefit from working with pupils of higher ability. Similarly, low ability pupils may absorb more classroom teacher contact time than higher ability pupils or disrupt teaching for other pupils by bad behaviour. Also, teacher allocation may be based on the makeup of the class. A school may allocate its best teachers to the lowest sets in order to maximise possible value added within the school, or similarly could allocate them to the highest sets in order to maximise the top level results possible.

4. Data

We use a unique sample from England consisting of 9,428 pupils taken in two tranches from a small sample of schools across the country. The data was collected at the Centre for Market and Public Organisation (CMPO) within the University of Bristol, for another purpose, namely to look at the effects of the introduction of teachers performance related pay in England (Atkinson et al. (2004)). Within the first tranche, we have 5,587 pupils within 35 schools who sat their key stage 3 examinations in 1997 and GCSE exams in 1999. Within the second tranche we have 3841 pupils within a subset of 23 schools who sat their key stage 3 examinations in 2000 and GCSE exams in 2002. These schools are a non-random sample of state schools, mixed sex and single sex, selective and non-selective. The sample was constructed purely on the basis that these schools were able and willing to divulge the extensive data requirements for the study aims. Hence, this dataset has the *unique* characteristic (for England) that we have complete data on the class in which all of the pupils are taught for English and maths. So, we have data on the pupil's entire classroom peer group, along with their abilities based on the key stage 3 scores already gained. We also have widespread but

---

[4] Lefgren (2004a) presents conditions that this IV strategy yields unbiased estimates of the peer effects and shows that the estimator is consistent when $\operatorname{cov}(KS3_{t-2}, s_t)^{UN} = \operatorname{cov}(KS3_{t-2}, s_t)^{TR}$, $\operatorname{var}(KS3_{t-2})^{UN} = \operatorname{var}(KS3_{t-2})^{TR}$ and $\operatorname{var}(s_t)^{UN} = \operatorname{var}(s_t)^{TR}$, where *UN* represents an untracked school (with R^2=0), and *TR* representing a tracked school, (with R^2=1).

incomplete knowledge of which teacher is taking each class. However, the pupil level data we observe is limited to age, gender and residential postcode.

Using postcode data on where each of the pupils lives we can map the pupils to a ward and thus can include data from the 2000 Indices of Deprivation from the Department of the Environment, Trade and Regions of local area deprivation. Hence, we have some idea of the demographics of the area in which the pupils live, and thus also some of the characteristics of their neighbours in terms of income, education, child deprivation etc.

In England, pupils sit compulsory key stage examinations at various points in their school career, at ages 7, 11, 14 and 16. At age 14, pupils are assessed at key stage 3 (KS3), a national examination, in English, mathematics and science, whilst at 16, pupils are assessed at general certificate of secondary education (GCSE) in a number of subjects, including English and mathematics. As a measure of ability we use the pupil’s key stage 3 and GCSE scores from English and mathematics. The key stage 3 scores are presented as a national curriculum level in the range from 2 to 8, and above that for exceptional performance. However, this exceptional level is very rare, and so we treat it as the same as those who receive a level 8 score. We also include an additional variable for those pupils who fail the key stage 3, or at least fail to gain a grade. The GCSE score is presented as a range from U (fail) to A*. In order to analyse the data, we consider an A* to be level 8 and a U to be a level 0.

We drop all results of pupils who are missing either a GCSE score or the subject specific key stage 3 score[5]. We also control for the age and gender of the pupils as well as including other ability measures consisting of the other subject key stage 3 scores.

GCSE qualifications are examined using a tier structure; with pupils being entered for the tier that the school decides is the best match to their ability. mathematics has three tiers; higher, intermediate and foundation, whilst English has two tiers; higher and foundation. Each tier only offers a range within the full grade spectrum. In English, a pupil can gain a grade in the range from A* to D for the higher tier paper and a range from C to G for the foundation paper. Similarly, for mathematics, a pupil can achieve a

[5] We thus drop results here for pupils classified with an X meaning entered but did not sit the exam

grade in the range A* to C for the higher tier, B to E for the Intermediate tier and D to G for the foundation tier. If a pupil fails at any tier they are awarded a U. Thus, a pupil of low ability entered for the foundation tier could receive an E grade whilst a higher ability pupil could be entered for the higher tier and fail, and thus receive a U, which could give the impression that the pupil entered for a lower tier paper had higher achievement. The nature of this tier structure further complicates the task of identifying the peer effects, since implicitly higher ability pupils will need to be taught to a higher syllabus to meet the requirements of the higher tier. Thus the content being taught is likely to be linked to the peer group. However, classes are not necessarily being taught a single tier. Some classes will have students taught as a mixed ability group with students entered for different tier exams at the end of the course.

In order to control for the different syllabus taught due to different tier entry, we need to control for this tier entry. We cannot directly observe what tier a pupil is entered for, but we can obtain an indicator as to what tier a set is collectively entered for based on the results gained at GCSE. It is a reasonable assumption that for many sets within schools the entire set will be entered for the same GCSE exam, since for each tier different syllabi are required. We examine the maximum and minimum scores pupils within the set achieve (excluding failures). We can subsequently compare this range with the range available within each tier, and if the results lie clearly within one tier, we assign that tier to the set. However, there is the potential for results not to point to one particular tier. For example, if in an English set, the only results gained were Cs and Ds we would not be able to distinguish between higher and foundation. We consider these sets where we cannot differentiate as being in the higher of the two possible tiers. This seems rational since in some of these borderline sets, whilst the top grades available in the lower tier are gained, some pupils also failed the exam. It is more likely that if the entire set were entered for the lower of the tiers, some of the lower grades would have also been obtained. However, this should not make a significant difference to the results, but we also examine the robustness of our results to assigning these borderline cases to the lower of the tiers.

For those sets where the exam results point to pupils entered in more than one tier within the set, we consider the set to be of mixed ability. For instance, if in mathematics, the maximum grade achieved within the set was an A* and the minimum mark was an E, the set could not all have been entered for higher tier or intermediate tier. We thus classify this as mixed set. This does not, however, distinguish between a high mixed ability set and a low mixed ability set in mathematics which has three tiers.

In order to construct the peer ability variable we consider the mean average of the key stage 3 scores of the other pupils within the class. Whilst at key stage 3, all pupils receive one grade in English, at GCSE; there is the possibility of receiving two GCSEs in English (language and literature). Having compared the structure of the English key stage 3 with the GCSEs, it was decided to use the mean average of the language and literature GCSE scores, with pupils who were missing either a language or literature score simply taking the non-missing score.

Our estimation method is within schools, utilising school fixed effects. Because of the way that we calculate our peer score, there will be a small within class variation. However, this is very small compared to the variation that is seen across classrooms. We thus only consider those schools where there is more than one class. Because of this, we lose a number of the schools that are small and only have one set for each subject. Similarly looking at within tier specifications, a larger number of schools will not have more than one set. Table 1 shows the number of schools that have a given number of sets both in the full sample and the restricted samples within tier entry for the set, and thus the number of schools that are included in our sample, once we have dropped those with less than one set.

Table 2 reports summary statistics for the pupils in our sample. The national average key stage 3 score for English is approximately 5, whilst for mathematics, the average score is just above 5. We have a slightly lower proportion of males than the national average of 0.511 in our sample for English and maths. The gender mix is not constant across the tiers with far fewer boys in the top tier and far more in the foundation tier, especially for English. Atkinson et al (2004) further discuss the representativeness of the sample of schools used in the study on a national level, and show that the “sample of schools is not, therefore, very representative of the national picture in terms of value added and GCSE scores.

Looking at the R-squared measure of setting, for whole schools, there is a relatively large value of 0.510 on average for English, and 0.749 for mathematics within schools. However, as discussed earlier, this is mainly due to the fact that within schools, the fact that GCSEs are examined in tiers, we would expect the R-squared for the whole school to be high compared to the R-squared setting measure for within tiers. This is evidenced further in table 3. The R-squared values for the within tier specifications are lower than those for the whole school, indicating a relatively less homogenous distribution of key stage 3 scores. That is, the lower R-squared measure within tier indicates a more random distribution of pupils to sets within the tier. This we can attribute to schools placing more emphasis on trying to ensure pupils are in a class teaching to the correct tier for GCSE. There is thus much more randomness when it comes to class allocation policies within the tier. It may be the case that for some schools there is a strict setting policy for within tier teaching whilst for others classes are taught in parallel with mixed ability within the class subject to being taught the appropriate tier. For these reasons, we may expect to see more robust results when we consider within tier results.

A worry is that the R-squared for mathematics is substantially larger than that for English in the higher tier, although again, as there are three tiers of entry in mathematics, and only a finite number of grades available at key stage 3, we would expect a more homogenous distribution of grades within tier in mathematics than in English. Our identification strategy assumes that schools with an R-squared of less than 0.35 will have a credibly random distribution of pupils by ability within the tiers. Figure 1 shows the distribution of the R-squared values within tier for English. We can see that for within whole schools, there are a wide range of setting policies, going from credibly random, with an R-squared of close to zero, to very strictly grouped according to ability, with an R-squared of close to 1. Within the higher tier, there is less variation in setting policy, but there is evidence of a considerable number of schools randomly assigning pupils into sets by ability, evidenced by the large proportion of schools with an R-squared value close to zero. In the foundation tier, there is evidence of more variation in the setting policies, with again, more apparently random setting policies within the foundation tier than within the entire school.

Figure 2 shows the distribution of the R-squared setting measure for mathematics. For the whole school case, it is immediately clear that there is much less heterogeneity of setting policies between schools, with the vast majority of schools having very strong policies regarding setting, evidenced by the large R-squared value. In the higher tier, there is evidence of more random sorting than in the whole school case, but there are still not many schools with very low R-squared measures indicating apparently random distribution of pupils. In the intermediate and foundation tiers, there is evidence of more schools having random setting policies than in the higher tier case, although also with more heterogeneity in setting policies across schools.

In our analysis, we use a measure of previous ability, the pupil's key stage 3 scores. The Key Stage 3 score is a national test sat by 14 year olds in English and maths. We also need to consider how to enter this prior achievement into our regressions as the effects may not be linear against the GCSE score. For all of the key stage 3 scores, we enter a failure as a separate dummy. This is due to the fact that as with GCSEs, the key stage 3 tests are examined in a tier structure with certain grades only available from certain tiers, and thus a failure is not necessarily representative of a child's ability[6]. Furthermore, we include all of the subject specific key stage 3 scores as individual dummies[7]. For other subject key stage 3 scores, we consider them to be linear between scores of 2 and 8, and similarly use a failure dummy to deal with the non-linearity we experience here.

5. Results

OLS Estimates

Table 4 contains OLS estimates of the classroom level peer effects present for English and maths. The regressions build up from a very simple model with no attempt to condition on prior attainment of the pupil concerned. This simply reflects the correlation between individuals' attainment and that of their peers conditional on the small set of demographic and deprivation indicators. Sequentially, the columns present regressions that include pupil prior attainment in the subject considered (column II) and in column III prior attainment in the other KS3 subject is also included. Column IV

[6] Mathematics is examined in 4 tiers, offering levels 2 to 5, 3 to 6, 4 to 7 and 5 to 8. English is examined in a single tier for reading and writing, the raw scores of which are added together to be converted into a national curriculum level.

[7] Upon testing linear effects of key stage 3 scores on GCSE scores in a regression of GCSE scores on a full set of score dummies for key stage 3, we reject the null of linearity for English at all reasonable significance levels (P>F=0.0000). We do not reject the null of linearity for mathematics, but for consistency we treat this in the same way as for English

introduces school fixed effects so that we are estimating within schools and finally Column V introduces teacher fixed effects. In Column V those teacher or teacher combinations that appear only once and retained in the sample and in Columns VI and VII we repeat columns IV and V but only include observations where the teacher is observed teaching at least two classes.

Pooled estimation

Starting with the English results in the upper panel of Table 4 the correlation between a pupil's attainment and his peers lagged attainment is very strong if we condition on only a limited range of personal, school and neighbourhood indicators. The coefficients imply that when the peer average lagged attainment changes by one grade[8], a result is seen equivalent to raising a pupil's attainment by 1 to 1.5 GCSE grade. The rows reflect the impact of moving to within tier estimation for English in the upper panel and maths in the lower. Within tier estimates are around 20-30% lower than for the full sample. The examination tiering is a major reason for setting and suggests that setting does create an upward bias to estimates of peer group effects.

Such models do not condition on pupils prior attainment, school intake selection or effectiveness or indeed teacher effectiveness. Introducing controls for the pupils' prior attainment (including any prior peer group effects) sharply reduces this correlation. Refining the prior attainment measure by including attainment in other KS3 subjects further reduces the correlation between pupil attainment and prior attainment of their classmates. The introduction of school fixed-year effects pushes the point estimate of the peer group effect upward and conditioning of teacher fixed effects makes no further difference. Restricting the sample to those pupils whose teachers are observed taking more than one class leaves the estimates unaffected, although due to the decreased sample size the standard error is increased. The introduction of school and teacher fixed effects within this relatively small sample of schools makes little difference to estimated peer group effects once pupil prior attainment is conditioned on as fully as possible. The estimates in columns IV and V suggest that an increase in average peer ability of one grade at key stage 3 in English raises pupil attainment by 0.4 GCSE

[8] This change is roughly a change of 1 standard deviations of the class average score in English, and 0.8 standard deviations for mathematics. Standard deviation of class average in English is 1.084, whilst for maths it is 1.243

grades or around one quarter of a standard deviation. Thus these estimated effects of peer effects within the classroom are moderately large. The picture for maths is broadly the same except that the estimated coefficients are somewhat higher with conditional estimates of around 0.6.

However, these estimates may still be misleading. As discussed earlier in the data section, in English schools there is a large amount of enforced stratification of pupils, due to the tiered nature of the examinations, so the highest ability students are never taught in a classroom at GCSE with the lowest ability students. It is thus more reliable to examine the effects within examination tier.

Within tier estimation

As noted earlier, at the school level setting is very common, especially in maths in order to, in part, facilitate teaching to a single exam tier. So when we consider within tier estimates the results are closer to a random allocation of pupils to classes, although there is a wide variation in school practices. The within tier estimates become very similar to whole sample estimates once we control for the child's past attainment as fully as we can. This suggests that including pupils' prior attainment captures the bias that setting for exam tiers produces or to put it another way the pupils KS3 scores provide the information used in grouping the children for entry into a GCSE exam tier.

For English the estimates without school or teacher fixed effects are smaller than in the full school regressions but the school fixed effect raises the estimates for an increase of the peer ability by 1 grade[9] at key stage 3 in higher and foundation tiers back to around 0.4.

For maths, the coefficient is higher for the higher tier, but this decreases as we move through intermediate, foundation to mixed tier teaching. A one grade increase in the class average at key stage 3[10] leads to an increase in individual pupil's attainment of

[9] This is equal to a change in the peer group of 1.5 standard deviations in higher tier, and 1.33 standard deviations in foundation tier. Standard deviation in higher tier English is 0.667 and foundation tier is 0.767.

[10] This is equal to a change in the peer group of 1.33 standard deviations in higher tier and approximately 2 standard deviations in intermediate and foundation tiers. Standard deviation in higher tier mathematics is 0.722, intermediate tier is 0.559 and foundation tier is 0.484.

approximately 0.76 grades in higher tier, 0.64 grades in intermediate and 0.4 grades in foundation tier.

Apparently random allocation of pupils.

The major concern is that despite within tier estimates, lagged pupil attainment, school and teacher fixed effects, there still be selection of pupils into classes within the school on the basis of unobserved (to the researcher but not the school) differences in pupils' ability leading to a possible bias in the estimates of the effect of an increase in the peer ability. Table 5 shows the results comparing the coefficients gained for the schools with low R-squared measures from within tiers with those that have a high R-squared measure. We now focus on the subset of schools that have a much lower R-squared setting measure, and thus a more credibly random distribution of the ability of pupils within the tier. In English, the picture is very clear cut. In both the higher and foundation tiers, the schools that have a low R-squared value, and consequently a credibly random distribution of pupils within the tier, have considerably lower estimates of the effect of a more able peer group than the OLS estimates on the full sample within each tier, whilst the schools with a high R-squared setting measure have considerably higher estimated effects than those seen in the full sample OLS regressions. For the higher tier estimation using credibly random distribution of pupils, there is a significant effect of a more able peer group demonstrated using our identification strategy, equivalent to an increase of between 0.17 and 0.20 grades for a one grade increase in the class average measure. For the foundation tier, a significant positive effect is seen, equivalent to an increase of between 0.23 and 0.28 grades for a one grade increase in the class average measure. This same picture is seen with the schools with apparently random setting policies within foundation tier mathematics, with a small positive effect observed in specification 4, although the significance is greatly reduced in specification 5, with smaller effects observed than in the OLS case. The magnitude of these effects is approximately an increase of between 0.230 and 0.296 grades for a one grade increase in the peer ability measure. Whilst these effects are smaller than those seen in the full sample, they are still positive, significant and non-trivial

As observed in the data section, the R-squared scores in the higher and intermediate tiers for mathematics are unaffected by concentrating on schools with apparently at random setting. The pattern that emerges is similar to that in Table 4 with higher estimates of peer effects in maths and especially higher tier maths and low estimates in English. However, the apparently at random setting schools for English suggest there was a moderately large upward bias to the estimates in Table 4. This may be because of there only being two rather than 3 exam tiers in English.

IV Estimation

It is possible that our selection of the "credibly random" sub-sample may still mask some underlying selection, leading to a residual bias of the estimate of the effects. In order to check the validity of our results, we use the identification strategy developed in Lefgren (2004b). Table 6 shows the first and second stage 2 stage least squares results, using the identification strategy developed by Lefgren (2004b).

The estimates within tier where there is far weaker evidence of active setting are very robust. An effect of similar size as seen in Table 5 is seen in English across the higher tier and foundation tier, meaning we cannot see any differential effect across ability ranges. However, by the very nature of the tiering, the lowest ability pupils are not placed with the highest ability pupils, and if they were, then we may expect to see a larger effect become apparent for the lower ability pupils. The estimated effect of a one grade change in the peer measure leads to approximately a one third of a GCSE grade, slightly lower than for the uninstrumented estimates in Table 4 but very similar to the apparently random sample seen in Table 5.

The estimates for maths only show significant effects for the intermediate and foundation tiers, and this becomes insignificant for foundation tier when we include the teacher fixed effects but the magnitude is very much in line with the estimate in Table 5. Within the intermediate tier, we see the strongest effect of having higher ability peers, with it actually increasing when we condition for the classroom teacher. This gives us an effect of about three tenths of a grade when moving one standard deviation in the peer measure. The estimated effects of peer group in higher tier maths are insignificant from zero and significantly different from the estimates in Table 5. This alternative approach produces result very much in line with our apparently random sample except for higher tier maths.

In order to test the endogeneity of the peer ability measure, we consider the OLS specification, but also include the residual obtained from the first stage of the two-stage IV regression. Table 7 shows the results of the endogeneity test. We can see that for English, the coefficient on the residual is not significantly different from zero for any of the within tier regressions, implying that the peer ability measure may not be endogenous. For maths, the story is more complicated, with the coefficient on the residual for the full sample being highly significant, but also there is significance on the higher tier and the mixed tiers, with a very low significance on the foundation tier. This difference in behaviour can be simply explained by recalling the summary statistics of the R-squared setting measure. For all tiers, the value was higher for maths than for English, implying that whilst there may be approximately random assignment of children, within tier, to classes in English, there is a more systematic policy for mathematics.

We may also wish to compare outcomes of studying in a class for foundation tier and higher tier. For this comparison, a school needs to have 2 or more sets of each tier. In order to make the marginal comparison, we consider sets as ordered by their average key stage 3 score, and compare the outcomes a borderline student would achieve in the highest foundation tier class and the lowest higher tier class (in the case of mathematics we consider the lowest intermediate tier class). For our comparison, we use specification IV, school fixed effects but not teacher fixed effects. This gives an average improvement of 0.66 grades by being in the higher tier classroom than in the lower tier classroom and for maths an average improvement of 0.62 grades.

6. Conclusions

We find significant and non-trivial evidence of peer effects within the classroom when both conditioning on school and teacher fixed effects. The examination system in England at GCSE with various different tiers encourages schools to teach children in sets grouped by ability in order to meet the differing requirements of the tiers. However, if we consider the grouping within the tier we find evidence of much more credible near random allocation within some schools.

We find very similar results using the sub-sample of schools that apparently allocate pupils (near) randomly within an exam tier and for the Instrumental Variables approach. These estimates of the effect of a more able peer group are approximately one fifth to one half of the unconditional OLS specifications and half to parity of those for conditional OLS estimates. Our OLS estimates on the schools with apparently random distribution of pupils give estimates of the effects that are not significantly different from the IV estimates, except for within higher tier mathematics and English. It is apparent from Figure 2 that within mathematics, there is a much higher level of setting in higher tier than in the other tiers, so there is a worry that the results may well be biased, and thus less robust than those for the IV specification. However, the IV estimates still give non-trivial significant effects for English, and for intermediate and foundation tiers for mathematics.

Our within tier teaching allows us to compare differential effects for pupils studying like exams, whilst pooled regressions may suffer from the fact that pupils are not necessarily studying the same syllabus and may be thus able to achieve differentially. In comparing pupils being taught in different tiers we see a considerable gap, which is difficult to attribute simply to being in a class with higher peers, and it may be necessary to attribute some of this gap to the difference in exam, and possible difference in aspirations due to being in a class where it is difficult to achieve even the most basic "pass" grade in GCSE. This is particularly important for the mathematics tiering as those in the foundation tier are pre-destined to be unable to reach the minimum level required to progress of a grade C. In fact Smith (2004) comments on the fact that nationally 30% of all pupils are pre-destined to fail GCSE maths before even sitting the exam simply due to the tier they are entered for. This may lead to low aspirations, and the carrot in intermediate tier of being able to gain a grade B could act to increase pupils' aspirations and thus increase their outcomes.

Whilst for each subject we see an improvement by being in the higher level classroom, there is still a question that remains of whether this is solely down to the influence of the peers, or whether this is more to do with the structure of the tiered examination. It may be of interest for further research to consider the effect that being entered into a higher tier examination has on the borderline children, especially those taught solely in a set being entered for the higher tier paper.

References

Angrist, J.D. and Lang, K (2004)., *Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program.* American Economic Review,. **94**(5):. 1613-34.

Atkinson, A., Burgess, S., Croxson, B. and Gregg, P., Propper, C., Slater, H., Wilson, D., (2004) *Evaluating the Impact of Performance-related Pay for Teachers in England*., CMPO Working Paper 04/113, CMPO, University of Bristol

Betts, J.R. and Shkolnik, J.L. (2000), *The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools.* Economics of Education Review,. **19**(1): 1-15.

Betts, J.R. and Zau, A. (2004), *Peer Groups and Academic Achievement: Panel Evidence from Administrative Data*, Public Policy Institute of California,

Bradley, S. and Taylor, J. (2006), *Do Peers Matter? Estimation of Peer Effects from Pupil Mobility Between Schools*, LERG discussion paper 13, CEPM, University of Queensland.

Burgess, S., McConnell, B., Propper, C., and Wilson, D.(2004), *Sorting and Choice in English Secondary Schools*. CMPO Working Paper 04/111, CMPO, University of Bristol,

Burke, M. and Sass, T.(2006), *Classroom Peer Effects and Student Achievement*. Working Paper, Florida State University,

Clark, C. (2003), *Measuring Peer effects with older peers: Sixth Grade Students in Middle School* Working Paper, University of Kentucky,

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., York, R.L. (1966), *Equality of Educational Opportunity*, US Department of Health, Education & Welfare, Office of Education.

Dills, A.K. (2005), *Does cream-skimming curdle the milk? A study of peer effects.* Economics of Education Review,. **24**(1): 19-28.

Gibbons, S. and Machin, S. (1998), *Valuing English Primary Schools.* Journal of Urban Economics, **53**(2): 197-219.

Gibbons, S. and Telhaj, S. (2006), *Peer Effects and Pupil Attainment: Evidence from Secondary School Transition*. Centre for the Economics of Education,

Hanushek, E.A., Kain, J.F., Markman. J.M., Rivkin, S.G. (2003), *Does peer ability affect student achievement?* Journal of Applied Econometrics,. **18**(5): 527-544.

Henderson, V., Mieszkowski, P. and Sauvageau, Y. (1978), *Peer Group Effects and Educational Production Functions.* Journal of Public Economics,. **10**(1): 97-106.

Hoxby, C. (2000), *Peer Effects in the Classroom: Learning from Gender and Race Variation*, NBER Working Papers: 7867,

Jencks, C. and Mayer, S.E. (1990), *The Social Consequences of Growing up in a Poor Neighbourhood*, in *Inner City Poverty in the United States*, Edited by L.E. Lynn, Jr., and M.G.H. McGeary, Nat. Acad. Press: Washington. ,

Lavy, V., Schlosser, A. (2007), *Mechanisms and Impacts of Gender Peer Effects at School*, NBER Working Papers: 13292,

Lefgren, L. (2004a), *Educational Peer Effects and the Chicago Public Schools*. Brigham Young University.,

Lefgren, L. (2004b), *Educational Peer Effects and the Chicago Public Schools.* Journal of Urban Economics,. **56**(2): 169-91.

Manski, C.F. (1993), *Identification of Endogenous Social Effects: The Reflection Problem.* Review of Economic Studies,. **60**(3): 531-42.

Mayer, S.E. (1991), *How much does a High School's Racial and Socioeconomic Mix Affect Graduation and Teenage Fertility rates?*, in *The Urban Underclass*, Edited by C. Jencks and P.E. Patterson, , Brookings Institution Press: Washington, DC.,: 79-102.

Proud, S. (2008) *Girl Power? An analysis of peer effects using exogenous changes in the gender make-up of the peer group* CMPO Working Paper 08/186, CMPO, University of Bristol.

Robertson, D. and Symons, J. (2003), *Do peer groups matter? Peer group versus schooling effects on academic attainment.* Economica,. **70**(277): 31-53.

Sacerdote, B. (2001), *Peer Effects with Random Assignment: Results for Dartmouth Roommates.* Quarterly Journal of Economics,. **116**(2): 681-704.

Smith, A. (2004), *Making Mathematics Count, the Report of Professor Adrian Smith's Inquiry into Post-14*, DfES.

Stinebrickner, R. and Stinebrickner, T.R. (2006), *What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds.* Journal of Public Economics,. **90**(8-9): 1435-54.

Summers, A.A. and Wolfe, B.L. (1977), *Do Schools Make a Difference?* American Economic Review,. **67**(4): 639-52.

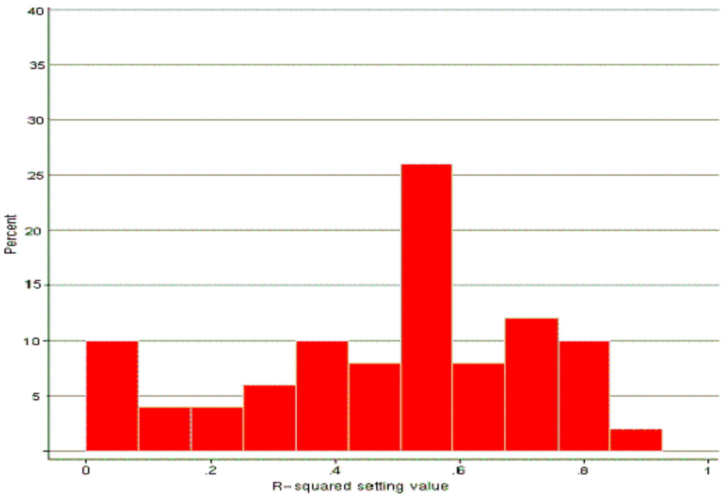
Todd, P.E. and Wolpin, K.I. (2003), *On the Specification and Estimation of the Production Function for Cognitive Achievement.* Economic Journal,. **113**(485): F3-33.

Winston, G.C. and Zimmerman, D.J. (2004), *Peer Effects in Higher Education*, in *College choices: The economics of where to go, when to go, and how to pay for it. NBER Conference Report series*, Edited by C.M. Hoxby, University of Chicago Press: Chicago and London: 395-421.

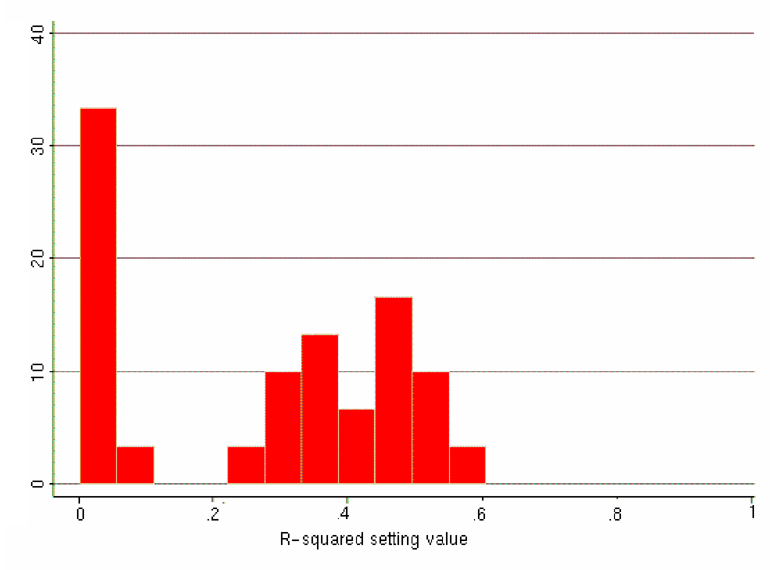
Zimmerman, D.J. (2003), *Peer Effects in Academic Outcomes: Evidence from a Natural Experiment.* Review of Economics and Statistics,. **85**(1): 9-23.

Figure 1 Distribution of R-squared setting measure for English between school

a) Whole Schools



b) Higher Tier



c) Foundation tier

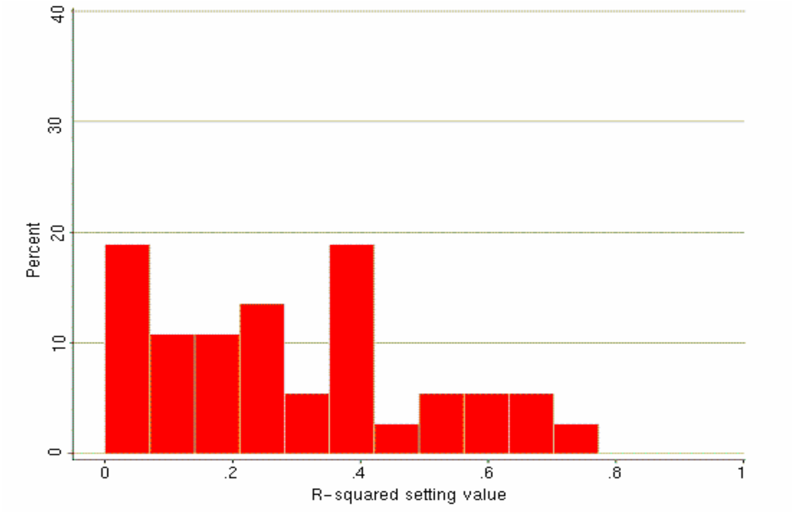
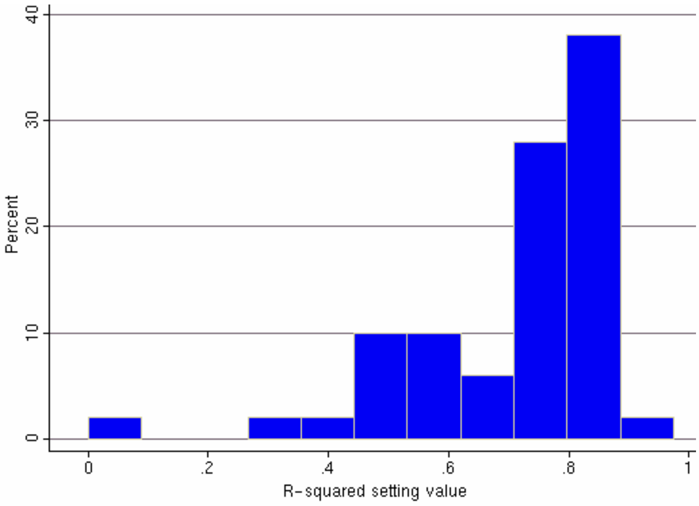
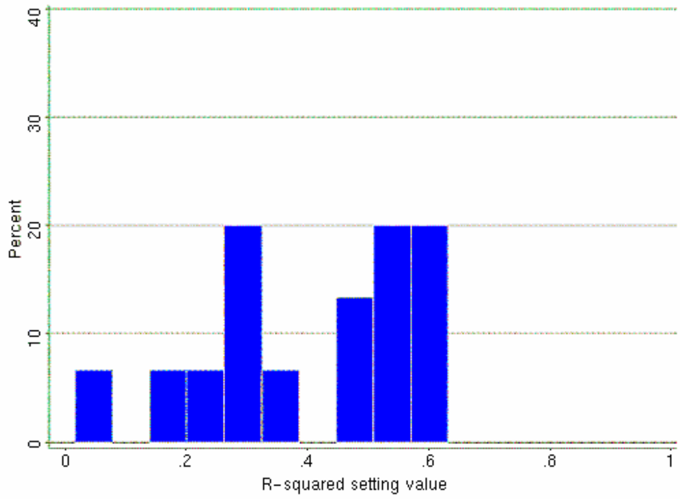


Figure 2 Distribution of R-squared setting measure for mathematics between schools

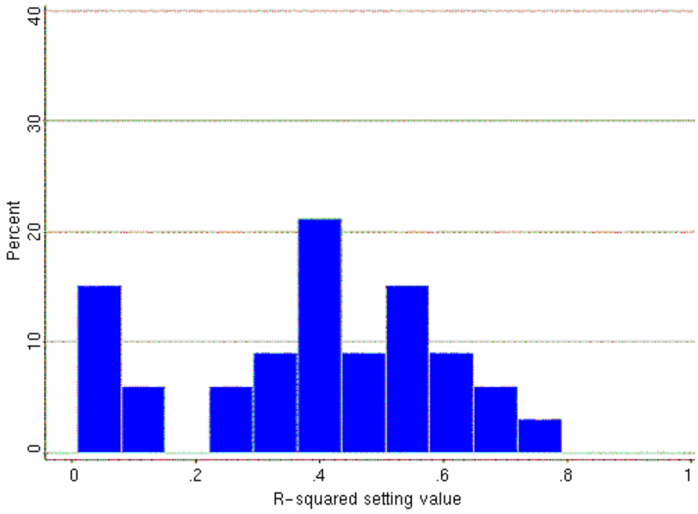
a) Whole Schools



b) Higher tier



c) Intermediate Tier



d) Foundation tier

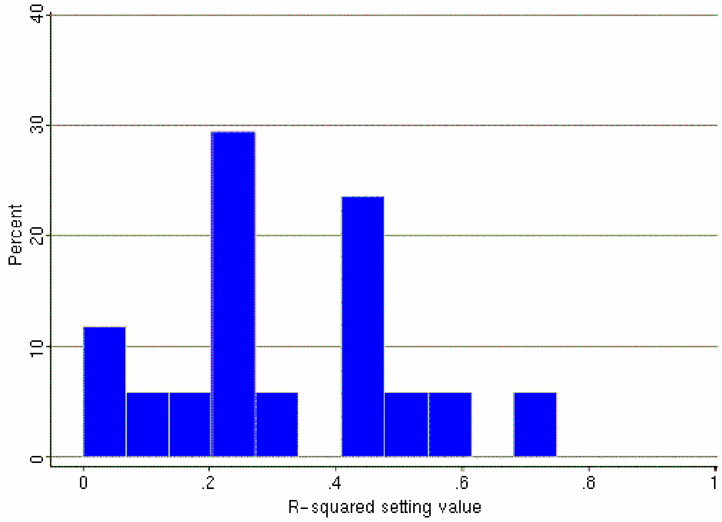


Table 1 Number of school/years with specified number of sets

No. Sets	Full Sample		Higher Tier		Inter-mediate Tier	Foundation Tier	
	English	Maths	English	Maths	Maths	English	Maths
1	2	1	13	21	13	5	9
2		1	13	7	24	12	16
3	2	1	8	5	3	12	16
4	8	8	4		4	10	5
5	12	9	4	2		1	
6	7	12		1	2	1	2
7	10	10					
8	6	3	1				
9	3	5				1	
10	2	3					
11							
12							
13							
14							
15	1						
16	1	2					
17							
18							

Notes: This table shows how many schools have each number of sets within each tier. The within tier regressions only consider schools with 2 or more sets within the tier.

Table 2 Summary statistics

Subject	GCSE Score	Key Stage 3 score	Age	Gender	R^2 setting measure	Sample size
Full Sample						
English	**4.756**	**5.061**	**16.259**	**0.508**	**0.510**	**6935**
	(1.559)	(1.155)	(0.296)	(0.500)	(0.197)	
Maths	**4.423**	**5.380**	**16.257**	**0.510**	**0.749**	**7231**
	(1.813)	(1.276)	(0.294)	(0.500)	(0.127)	
Higher tier						
English	**5.824**	**5.792**	**16.276**	**0.465**	**0.263**	**2328**
	(1.034)	(0.919)	(0.291)	(0.499)	(0.203)	
Maths	**6.406**	**6.815**	**16.257**	**0.496**	**0.443**	**1170**
	(0.941)	(0.718)	(0.297)	(0.500)	(0.171)	
Schools with low R-squared measures						
English	**5.961**	**5.899**	**16.270**	**0.370**	**0.044**	**987**
	(1.029)	(0.990)	(0.285)	(0.483)	(0.065)	
Maths	**6.264**	**6.744**	**16.255**	**0.555**	**0.278**	**523**
	(0.955)	(0.713)	(0.300)	(0.497)	(0.113)	
Schools with high R-squared measures						
English	**5.859**	**5.693**	**16.285**	**0.547**	**0.470**	**909**
	(1.070)	(0.864)	(0.300)	(0.498)	(0.059)	
Maths	**6.742**	**6.942**	**16.258**	**0.350**	**0.605**	**446**
	(0.834)	(0.697)	(0.292)	(0.477)	(0.032)	
Intermediate tier						
Maths	**4.567**	**5.466**	**16.258**	**0.507**	**0.401**	**2030**
	(1.016)	(0.705)	(0.297)	(0.500)	(0.207)	
Schools with low R-squared measures						
	4.669	**5.571**	**16.267**	**0.548**	**0.198**	**834**
	(1.029)	(0.709)	(0.291)	(0.498)	(0.153)	
Schools with high R-squared measures						
	4.450	**5.410**	**16.253**	**0.475**	**0.594**	**786**
	(1.046)	(0.641)	(0.294)	(0.500)	(0.063)	
Foundation tier						
English	**3.140**	**4.028**	**16.232**	**0.640**	**0.313**	**1724**
	(1.097)	(0.975)	(0.303)	(0.480)	(0.200)	
Maths	**2.291**	**3.851**	**16.233**	**0.552**	**0.390**	**1521**
	(1.050)	(0.674)	(0.289)	(0.497)	(0.179)	
Schools with low R-squared measures						
	3.048	**3.917**	**16.234**	**0.613**	**0.116**	**686**
	(1.135)	(1.071)	(0.299)	(0.487)	(0.083)	
	2.226	**3.756**	**16.231**	**0.518**	**0.196**	**606**
	(1.022)	(0.643)	(0.277)	(0.500)	(0.093)	
Schools with high R-squared measures						
	3.113	**4.128**	**16.239**	**0.706**	**0.517**	**656**
	(1.063)	(0.806)	(0.313)	(0.456)	(0.122)	
	2.465	**3.981**	**16.239**	**0.635**	**0.569**	**572**
	(1.054)	(0.704)	(0.297)	(0.482)	(0.058)	

Note. Standard deviations in parentheses. Unit of observiation is an individual child.

Table 3 Description of regression specifications.

Specification	Description
I	We include age, gender, index of income deprivation and the proportion of pupils in the school who are male and a dummy for whether the school year has more than the mean number in it, indicating a large school.
II	Includes the subject specific key stage 3 score
III	Includes the other subject key stage 3 scores
IV	Includes school fixed effects
V	Includes teacher fixed effects (Teachers who teach 2 or more classes and all others including those identified as teaching 1 class in sample replaced as missing)
VI	Subsample of IV with identifiers for teachers who teach 2 or more classes
VII	Only with teachers who teach 2 or more classes. (Missings and teachers who teach only one class omitted)

Table 4 Results from ordinary least squares estimation of the effect of peer ability on outcomes.

	I	II	III	IV	V	VI	VII
School/year fixed effects				√	√	√	√
Teacher fixed effects					√		√
English							
Full Sample							
Class Average peer measure	**1.169***** (0.031)	**0.558***** (0.036)	**0.336***** (0.032)	**0.439***** (0.033)	**0.442***** (0.032)	**0.437***** (0.044)	**0.425***** (0.041)
Observations	6935	6935	6935	6935	6935	3776	3776
R-squared	0.53	0.62	0.69	0.72	0.74	0.74	0.76
Higher tier							
Class Average peer measure	**0.854***** (0.071)	**0.412***** (0.073)	**0.248***** (0.065)	**0.442***** (0.066)	**0.447***** (0.070)	**0.761***** (0.174)	**0.862***** (0.173)
Observations	2328	2328	2328	2328	2328	489	489
R-squared	0.34	0.43	0.53	0.64	0.64	0.75	0.75
Foundation Tier							
Class Average peer measure	**0.669***** (0.064)	**0.305***** (0.068)	**0.224***** (0.063)	**0.367***** (0.055)	**0.435***** (0.063)	**0.357***** (0.123)	**0.238** (0.146)
Observations	1724	1724	1724	1724	1724	420	420
R-squared	0.20	0.28	0.36	0.43	0.44	0.50	0.51
Mathematics							
Full Sample							
Class Average peer measure	**1.303***** (0.021)	**0.676***** (0.046)	**0.555***** (0.047)	**0.605***** (0.045)	**0.595***** (0.045)	**0.632***** (0.065)	**0.613***** (0.066)
Observations	7231	7231	7231	7231	7231	3675	3675
R-squared	0.70	0.74	0.75	0.79	0.80	0.81	0.82
Higher tier							
Class Average peer measure	**1.092***** (0.088)	**0.699***** (0.117)	**0.571***** (0.113)	**0.758***** (0.079)	**0.767***** (0.070)	**0.884***** (0.145)	**0.919***** (0.100)
Observations	1170	1170	1170	1170	1170	208	208
R-squared	0.39	0.44	0.47	0.58	0.60	0.62	0.62
Intermediate Tier							
Class Average peer measure	**0.982***** (0.084)	**0.542***** (0.090)	**0.441***** (0.088)	**0.630***** (0.081)	**0.650***** (0.083)	**0.728**** (0.264)	**1.055**** (0.347)
Observations	2030	2030	2030	2030	2030	313	313
R-squared	0.26	0.30	0.33	0.43	0.43	0.41	0.43
Foundation Tier							
Class Average peer measure	**1.045***** (0.085)	**0.502***** (0.092)	**0.375***** (0.092)	**0.457***** (0.070)	**0.400***** (0.065)	**0.926***** (0.273)	**0.894***** (0.291)
Observations	1521	1521	1521	1521	1521	208	208
R-squared	0.28	0.37	0.39	0.49	0.51	0.52	0.52

Notes Dependent variable is the GCSE score in English or mathematics. Specifications of regressions shown in table 3. Method of estimation is ordinary least squares. (OLS) Robust standard errors for within class clustering in parentheses. * indicates significant at 10%; ** indicates significant at 5%; *** indicates significant at 1%

Table 5 Results from the estimation of the effect of a more able peer group using schools that have a credibly random distribution of pupils by ability within tiers.

	English		Maths	
	IV	V	IV	V
School/year fixed effects Teacher fixed effects	√	√ √	√	√ √
1. Higher Tier **OLS Low R-squared** Class Average peer measure	0.198** (0.082)	0.167** (0.069)	0.820*** (0.182)	0.806*** (0.113)
Observations R-squared	1330 0.67	1330 0.67	469 0.43	469 0.47
High R-squared Class Average peer measure	0.524*** (0.097)	0.518*** (0.118)	0.773*** (0.115)	0.792*** (0.109)
Observations R-squared	770 0.63	770 0.63	701 0.66	701 0.67
2. Intermediate Tier **Low R-squared** Class Average peer measure			0.772*** (0.139)	0.769*** (0.184)
Observations R-squared			633 0.47	633 0.48
High R-squared Class Average peer measure			0.626*** (0.109)	0.634*** (0.112)
Observations R-squared			1144 0.42	1144 0.43
3. Foundation tier **Low R-squared** Class Average peer measure	0.291*** (0.090)	0.331*** (0.099)	0.296** (0.143)	0.230* (0.135)
Observations R-squared	936 0.39	936 0.40	588 0.44	588 0.44
High R-squared Class Average peer measure	0.446*** (0.091)	0.545*** (0.098)	0.556*** (0.079)	0.512*** (0.075)
Observations R-squared	556 0.46	556 0.47	933 0.54	933 0.56

Notes Dependent variable is the GCSE score in English or mathematics. Specifications of regressions shown in table 3. Method of estimation is ordinary least squares. (OLS) Robust standard errors for within class clustering in parentheses. * indicates significant at 10%; ** indicates significant at 5%; *** indicates significant at 1%. Low R-squared indicates a school with an R-squared score less than 0.35. High R-squared indicates a school with an R-squared higher than 0.40.

Table 6 Results from two stage least squares estimation of the effect of peer ability on outcomes.

	English		Maths	
	IV	**V**	**IV**	**V**
School/year fixed effects	√	√	√	√
Teacher fixed effects		√		√
First stage of 2 stage least squares				
Higher Tier				
Higher tier instrument	**0.968*****	**0.931*****	**0.966*****	**0.951*****
	(0.082)	(0.091)	(0.160)	(0.160)
Observations	2328	2328	1170	1170
R-squared	0.80	0.81	0.72	0.72
Intermediate Tier				
Intermediate tier instrument			**0.926*****	**0.912*****
			(0.095)	(0.090)
Observations			2030	2030
R-squared			0.72	0.73
Foundation Tier				
Foundation tier instrument	**0.974*****	**0.752*****	**1.008*****	**0.935*****
	(0.110)	(0.160)	(0.109)	(0.117)
Observations	1724	1724	1521	1521
R-squared	0.67	0.76	0.62	0.66
Second stage of 2 stage least squared				
Higher Tier				
Class Average peer measure	**0.377*****	**0.380*****	**0.249**	**0.201**
	(0.126)	(0.133)	(0.229)	(0.206)
Observations	2328	2328	1170	1170
R-squared	0.64	0.64	0.56	0.57
Intermediate Tier				
Class Average peer measure			**0.581*****	**0.671*****
			(0.214)	(0.210)
Observations			2030	2030
R-squared			0.43	0.43
Foundation tier				
Class Average peer measure	**0.309*****	**0.309***	**0.304***	**0.266**
	(0.115)	(0.168)	(0.153)	(0.171)
Observations	1724	1724	1521	1521
R-squared	0.43	0.44	0.49	0.51

Notes Dependent variable is the GCSE score in English or mathematics. Specifications of regressions shown in table 3. Method of estimation is two stage least squares. Robust standard errors for within class clustering in parentheses. * indicates significant at 10%; ** indicates significant at 5%; *** indicates significant at 1%.

Table 7 Test of endogeneity of class peer ability measure

	English		Maths	
	IV	V	IV	V
School/year fixed effects	√	√	√	√
Teacher fixed effects		√		√
Higher Tier				
Class Average peer measure	**0.377***** (0.128)	**0.380***** (0.136)	**0.249** (0.225)	**0.201** (0.196)
Residuals	**0.079** (0.130)	**0.080** (0.137)	**0.566**** (0.278)	**0.627**** (0.233)
Observations	2328	2328	1170	1170
R-squared	0.64	0.64	0.59	0.60
Intermediate Tier				
Class Average peer measure			**0.581***** (0.214)	**0.671***** (0.210)
Residuals			**0.056** (0.217)	**-0.024** (0.217)
Observations			2030	2030
R-squared			0.43	0.43
Foundation Tier				
Class Average peer measure	**0.309***** (0.115)	**0.309*** (0.164)	**0.304*** (0.156)	**0.266** (0.173)
Residuals	**0.068** (0.127)	**0.141** (0.174)	**0.175** (0.167)	**0.152** (0.184)
Observations	1724	1724	1521	1521
R-squared	0.43	0.44	0.50	0.51

Notes Dependent variable is the GCSE score in English or mathematics. Specifications of regressions shown in table 3. Method of estimation is ordinary least squares. (OLS) Robust standard errors for within class clustering in parentheses. * indicates significant at 10%; ** indicates significant at 5%; *** indicates significant at 1%.