Author:
Proud, Steven

Title:
Evidence of Peer Effects in English Schools

Evidence of Peer Effects in English Schools

by

Steven Proud

A dissertation submitted to the University of Bristol in accordance with the requirements for award of degree of PhD in the Faculty of Social Sciences.

Department of Economics
September 2009

Word Count: 48,124

Abstract

The purpose of this thesis is to consider how a child's peer groups affect their educational outcomes up to the age of 16.

In the first empirical chapter, I estimate the effect of the gender mix on pupils' outcomes in national tests at ages 7, 11, 14 and 16 using exogenous changes within school in the proportion of the peer group that is female. The results suggest that boys perform significantly worse in English when the proportion of girls is increased, whilst boys and girls perform better in primary schools in mathematics and science with an increase in the proportion of girls.

The second and third empirical chapters offer estimates of the effect of a more able peer group on outcomes at ages 16 and 11 respectively using different methodologies. The former considers schools that have a credibly random allocation of pupils into classrooms within tiers for GCSE entry and uses an IV strategy to validate the results gained for the credibly random distribution. The latter uses the proportion of pupils who are old and young within the cohort as instruments for pupils' prior outcomes. Both these chapters suggest that the effect of a more able peer group is both positive and non-trivial.

Acknowledgements.

I would like to thank my supervisors, Professor Simon Burgess and Dr Deborah Wilson for the time and effort they have spent in the course of my studies. Also, thanks must go to Professor Paul Gregg, Professor Carol Proper and Adele Atkinson who were collaborators, along with Professor Simon Burgess in the writing of chapter 5. Thanks must go to the Economic and Social Research Council who provided the funding for this research and to the Department for Children, Schools and Families for providing the PLASC and NPD dataset. Also, to all my colleagues at the Centre for Market and Public Organisation and the University of Bristol and seminar participants for their helpful comments and suggestions.

Final thanks must go to my (more literate) friends and family who have provided great assistance in proof reading.

Author's Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE:..........................

Contents

Index of figures

Chapter 2

Chapter 4

Chapter 5

Chapter 6

Index of tables

Chapter 2

Chapter 4

Chapter 5

Chapter 6

Chapter 1 Introduction

"Education, education, education". This was the list of priorities spelled out by the Labour party in its push for election to government in 1997. McIntosh (2006) demonstrates that the premium for a good education is high, and suggests that gaining 5 or more good qualifications at age 16[1] gives a wage premium of 28% over gaining no good qualifications. But, what factors help to mould the educational outcomes of an individual?

A good place to start might be to consider the seminal work for sociologists and educationalists, the Coleman Report (Coleman et al (1966)), which was commissioned to examine the equality of educational opportunity for white and black pupils in the USA. Coleman et al (1966) comment that whilst there are differences in composition of the schools, the vast majority of total variation is explained by within school variation in pupils' achievement, the size of which dwarfs the variance between schools. These assertions are also considered in works such as Feinstein and Symons (1999), which suggest that parental input, most notably parental interest in education, has a much larger effect on outcomes than any school based input. Whilst these studies do place the majority of the variation in test scores within schools, Chevalier et al (2005) provide a review of the relevant literature, and show that there is a differential effect of both teachers and schools indicating there is room for choice from parents to try to maximise the outcomes that their child can gain.

Numerous models of the educational production functions, which tend to concentrate on individual, family and school characteristics have been proposed. Herbert and Thomas (1998) try to formalise a model by identifying key inputs into an individual's outcomes. Pupils have ability, which is a factor of their underlying aptitude and IQ. This underlying ability, when combined with their age, gender and other factors makes up the individual component of their prior outcomes. Added to this are more aggregated measures, including the size of the child's family, social class and neighbourhood factors. These underlying factors

[1] A good qualification at age 16 is defined as gaining a grade A* to C in General Certificate of Secondary Education (GCSE) examination, or equivalent.

help to create a level of underlying ability, which as discussed in Todd and Wolpin (2003) is also a function of any previous school inputs, and then this underlying ability is acted upon by a black box of school factors. Rivkin et al (2005) move further, and explicitly disentangle school and teacher effects from the black box of school factors. However, what are the factors within a school that influence (and hopefully maximise) children's outcomes?

The Coleman Report (Coleman et al (1966)) also introduces the possibility that a pupil's peers have a beneficial effect on an individual pupil's outcomes as it suggests that "a pupil's achievement is strongly related to the educational backgrounds and aspirations of the other students in the school" (Coleman et al (1966, 22)). They rank the order of influence on pupils outcomes, from most important to least, as pupil background, followed by the pupil's peer group and finally the school facilities and teaching.

Whilst Coleman et al (1966) identified the correlation between the outcomes of individuals and peer characteristics, it wasn't until Manski (1993) that problems inherent with estimating sorts of production functions including the effects of a child's peers became widely appreciated. Families sort themselves into neighbourhoods, whether by house price selection (Gibbons and Machin (2003)), or by some other factor, but as discussed in Burgess et al (2007) there is both residential and post-residential sorting of pupils into schools, with peers that more closely match their own attributes. Interestingly, Burgess et al (2007) and Urquiola (2005) also show that having conditioned on sorting into neighbourhoods, the level of post-residential sorting is increased with the availability of more choice.

Kramarz et al (2009) try to quantify the components that go into making an individual's test scores, and consider individual level factors, school factors and the effect of the peer group on outcomes. Their results suggest that the order of importance of influences on individual outcomes is, from most important, pupil background, then school factors and finally a small, but significant, effect of the peer group.

In this thesis, I examine the effect of a child's within school and within classroom peer groups on their academic outcomes within 3 empirical chapters. This follows inspiration from the Coleman Report (Coleman et al (1966)), and investigates the effect of the composition of a child's peer group on their outcomes at ages 7, 11, 14 and 16. In examining the effect of the composition of the peer group, I consider two distinct areas. First, I consider the effect of the gender composition of the peer group. That is, the effect of having a higher proportion of girls within the peer group on outcomes. The second distinct area I examine is the effect of a more able peer group on outcomes at ages 11 and 16.

The first empirical chapter, "Girl Power" investigates the effect of a more female peer group on pupils' outcomes at ages 7, 11, 14 and 16, and uses year on year changes within a school of the proportion of the peer group that is female to estimate the effect of a change in the gender make-up of the peer group. The results for all schools and levels suggest that a 10 percent increase in the proportion of the peer group that is female is associated with a 0.015 standard deviation *decrease* in boys' outcomes in English, whilst a 10 percent increase in the proportion of the peer group that is female leads to an increase in girls achievement in primary schools in mathematics and science. Further to this, in primary schools, there is a strong negative effect on boys in English of a more female peer group, but in mathematics and science, there is a strong positive effect for both boys and girls. As an innovation, this chapter takes advantage of legislation in English schools, which limits infant schools to a maximum class size of 30. I show that for examinations sat at age 7 and age 11, it is likely that schools that have 30 or fewer pupils in every observed cohort can be considered to have just 1 class per academic year, and as such, in primary school[2] I can estimate the classroom level effect of the peer group on pupils' outcomes. These results again suggest a significant negative effect of a more female peer group on boys' outcomes at key stage 1 in English, and a significant positive effect of a more female peer group on girl's outcomes in mathematics.

[2]Primary school covers pupils from age 4 to age 11.

The second empirical chapter, "Peer effects in English Secondary schools"[3] examines the effect of a more able peer group at the classroom level at General Certificates of Secondary Education (GCSE) level in England. This chapter uses a dataset which identifies the teachers and the classes which students are taught in from ages 14 to 16, and as such can identify exactly an individual child's classroom peer group. This chapter takes advantage of this dataset to estimate peer effects using two related methodologies. Pupils are entered for GCSE examination within tiers, in which only a subset of the total grades are available. Most schools teach their pupils in classes focussed on entry into one of these tiers, which leads at a school level to relatively homogenous classrooms. However, there is variation between schools in the level of homogeneity of pupils' prior ability within these tiered classrooms, varying from very strict setting policies to a more random allocation of pupils to classes within the tier. The first methodology estimates a measure of the level of sorting within the tier of entry for GCSE within a school to estimate which schools offer a credibly random distribution of pupils by ability to classrooms within the tier, and then uses this subset of schools to estimate the effect of a peer group using ordinary least squares (OLS). The second methodology, used to validate the results from the first methodology, again takes advantage of this measure of within school sorting, but uses this, interacted with a child's own ability to construct an instrument for the ability of the peer group, as developed by Lefgren (2004b). The estimates gained from the credibly random subset are larger than those seen in some other more recent studies, and suggest a one standard deviation increase in the ability of the peer group is associated with a between 0.07 and 0.22 standard deviation increase in individuals outcomes. The instrumental variables (IV) results are of a similar magnitude as those gained for the random sample, and all the results are significant and non-trivial.

The third empirical chapter, "Peer effects in English Primary schools" is a companion to the previous chapter. In this chapter, I use a new instrument to estimate the effect of a more able peer group within English primary schools. This chapter takes advantage of the correlation between the month of birth of an

[3] This chapter is co-authored with Adele Atkinson and Professors Carol Propper, Paul Gregg, and Simon Burgess.

individual child and their outcomes to create instruments for identification, namely the proportion of pupils within the peer group who are in the oldest third of the age distribution and the proportion of pupils within the peer group who are in the youngest third of the age distribution, and as with the first empirical chapter attempts to estimate the effect on a classroom level by considering schools that have 30 or fewer pupils registered in each wave of the dataset. The results suggest that a one standard deviation increase in the ability of the peer group is associated with a between 0.05 and 0.4 standard deviation increase in the outcomes of an individual child. The magnitude of the effects estimated in primary schools is largely similar to those estimated in secondary schools in the previous chapter.

The structure of this thesis is as follows. Chapter 2 discusses the structure of schools in England. Also examined in chapter 2 are the pupil level annual school census (PLASC) and the national pupil database (NPD) which are used as the data source for chapters 4, 6 and 7. Chapter 3 examines the previous literature on the estimation and problems inherent with estimating the effect of a more able peer group on children's outcomes. Chapter 4 examines the effect of a change in the gender-mix of a child's peer group. Chapter 5 takes advantage of a unique dataset to estimate the effect of a more able peer group at age 16 using both a credibly random distribution of pupils within a school and an instrumental variables methodology. Chapter 6 examines the effect of a more able peer group at age 11 using an IV strategy that takes advantage of variations in the age of a child's peer group. Finally, chapter 7 offers some concluding remarks.

Chapter 2 English Schools and the PLASC/NPD Dataset

2.1 Introduction

In chapters 4, 5 and 7, I use data from the National Pupil Database (NPD) and the Pupil Level Annual School Census (PLASC), containing data on all pupils in state funded education in England and their schools' characteristics. In this section, I begin by discussing the types of schools in England, followed by the structure of schooling and examinations within England. I then move on to examine the data that is contained within the NPD and PLASC. In section 2.5, I discuss decisions made regarding the sample selection from the population. In section 2.6, I examine a property of infant schools in England, which allows me to try to consider which pupils are educated with which. This section considers schools that have fewer than 30 pupils in them, and discusses why they can credibly be considered as consisting of only one class per school-year. Finally, in section 2.7, I examine selected summary statistics of the data.

2.2 Types of schools in England

In England, there are a number of types of state-maintained schools, which have different levels of control and ownership from the LA and the state[4]. Within state-funded schools, there are also specialist schools, which offer specialisation within a particular subject area, which could involve language, sports, music etc. In this thesis, I do not consider these specialist schools to be different from their non-specialist counterparts, as whilst they have a specialism in a particular area, they are still required to follow the same national curriculum that all other state-funded schools also follow, although specialist schools can select up to ten percent of their pupils based on ability within the school's particular specialism. Other than state funded schools, there are independent schools, which I do not consider here.

---

[4] For more details, see http://www.direct.gov.uk/en/Parents/Schoolslearninganddevelopment/ChoosingASchool/DG_4016312 , Department for Children, Schools and Families., accessed 10/8/2009

2.2.1 State Funded schools

Firstly, I consider schools which are funded by the LA. Community schools are schools which are funded by the LA, which also employs the staff and owns the land and buildings, and makes up the majority of maintained schools in England. Voluntary controlled schools are schools for which, whilst they are owned by a charity, the control of the school is the responsibility of the LA, rather than the school's own governing body. For both these types of schools, the LA decides on the criteria that the school uses for admissions, which may include the distance pupils live away from the school or whether they have siblings at the school.

Foundation and trust schools[5] are schools which are run by their own governing body, rather than the LA, who make crucial decisions, whilst the land and buildings are owned by the governing body, or an external charitable foundation. Voluntary-aided schools, which are often faith schools, are similar to foundation schools, since the governing body sets policy within the school, such as the admissions policy, and is also responsible for employing the staff. Again, the buildings are owned by a charitable foundation, which is often a religious organisation, such as the Church of England, whilst the governing body also contributes towards buildings and maintenance costs. These voluntary aided schools may have within school policies, such as requirements of admissions and teacher recruitment, which are consistent with the faith status of the school. For voluntary aided and foundation schools, the governing bodies set the school's oversubscription criteria, within certain rules. There are also a small number of city technology colleges which are independent of the state, but the running costs are funded by the state and pupils do not pay fees. These schools have one fifth of the capital cost paid for by private sponsors, and have different admissions policies to those in the rest of the state maintained sector. Pupils must live in the catchment area, and the intakes are representative of the ability range of pupils within the catchment area.

---

[5] Trust schools were introduced in 2007, but are not included in the data I use here.

Academies[6] are schools which are managed independently and are open to all abilities of pupils, although, as with specialist schools, they can select up to ten percent of their pupils based on the academy's specialism. Academies are set up by sponsors, such as churches, businesses or other groups, but in a partnership with the government and LA. Grammar schools are selective schools which select pupils based on ability, assessed usually in an examination at age 11. There are also a small number of maintained boarding schools, which offer free education to pupils, but charge a fee for pupils for board and lodging. Finally, there are a number of community special and foundation special schools, which cater solely to pupils with special education needs.

2.3 Structure of schooling in England

In England[7], it is compulsory for children to attend full-time education until the age of 16[8]. In most local authorities (LAs), pupils are educated in infant schools (from the September after they turn 4 until the September after they turn 7), junior school (from age 7 to 11) and secondary schools (from 11 onwards), and for many schools, the infant and junior schools are on one site, making up a single primary school. A small minority of local authorities educate their children in middle schools, from age 9 to age 13. Infant and junior schools tend to be significantly smaller than secondary schools, with several primary schools often acting as feeders to one much larger secondary school. In 2006, there were a total of 17,504 non-special state maintained primary schools and 3,367 non-special state maintained secondary schools in England[9], although these figures also include selective schools, academies and other schools not used in this thesis.

---

[6] Academies are not included in this thesis, as they are a new innovation, and no pupils results appear in the initial 2002 release of PLASC

[7] I only consider English schools here as the NPD dataset only contains data for English schools.

[8] This school leaving-age will be increased to age 17 in 2013 and 18 in 2015, following the Education and Skills act 2008.

[9] Numbers of schools available from http://www.dcsf.gov.uk/rsgateway/DB/TIM/m002003/index.shtml , accessed 21/09/2009

Children in schools run by the state are educated following a set national curriculum, which is split into key stages, assessed at ages 7, 11, 14 and 16 at the end of, respectively, key stage 1, key stage 2, key stage 3[10] and key stage 4. Key stage 4 is often examined as a set of General Certificates in Secondary Education (GCSEs), although some pupils are educated in vocational subjects, including the nationally set National Vocational Qualifications (NVQs). At key stage 1, pupils are assessed using tasks and tests in reading, writing and mathematics. Teachers use these tasks and tests, along with observations of speaking and science ability to give a level of achievement in reading, writing, speaking, mathematics and science. This teacher assessment is moderated by the LA [11]. At key stages 2 and 3, up until 2008, pupils are assessed by examinations in English, mathematics and science. Post 2008, pupils continue to be assessed by examination at the end of key stage 2, but at key stage 3, teachers are now required to assess the levels achieved by children in both the core subjects (English, mathematics, science) and the non-core subjects (history, geography, modern foreign languages, design and technology, information and communications technology, art and design, music, physical education, citizenship and religious education). At GCSE, pupils are assessed in a number of subjects, including English, mathematics and science, but are also examined in a choice of other subjects. The structure of schools, and the ages at which pupils sit examinations is summarized in Figure 1.

Pupils' achievement at key stage1, key stage 2 and key stage 3 is measured in national levels. In each subject, the national curriculum is separated into strands which assess various skills within the subject, and each level is associated with a certain skill level that needs to be achieved. Levels can be achieved between 1 and 8, with a further grade only available for exceptional performance. These are converted into national curriculum scores.

Within GCSE results are presented using the range of A* to G, with a U for a fail, A* indicating the highest grade and G the lowest. Whilst the GCSE grades are measured in a different way from the key stage levels, in order to quantify

[10]Current requirements for key stage assessment are available here: http://www.direct.gov.uk/en/Parents/Schoolslearninganddevelopment/ExamsTestsAndTheCurriculum/DG_10013041 accessed 02/09/2009

[11] In this thesis, I only consider reading, writing and mathematics at key stage 1.

the results, I use the national curriculum scores, with an A* being worth 58 points with each grade lower being worth 6 points less.

For GCSE and key stage 3, pupils are entered for specific tiers, in which the pupils cannot access all grades. For example, in English GCSE, pupils could be entered for the higher tier examination, and would only be eligible for grades A*-D, whilst if they were entered for the foundation tier, they could access the C-G grades[12]. In mathematics, a three tier system was in place for GCSE, but was phased out in favour of a two tier system by 2007. At key stage 3, there are a large number of tiers, but the difference between the tiers is less marked than at GCSE.

Key stage 3 examinations were discontinued in England following public outcry of the marking organisation and standards in 2008 and have been replaced with teacher assessment, as described above, but prior to this they were examined in tiers as with GCSE. For mathematics and science, tiers were available which offered grades 3-5, 4-6, 5-7 and 6-8. For English, all pupils were assessed automatically in the range 4-7.

2.4 Contents of the NPD and PLASC

The National Pupil Database (NPD) is a database that came into existence in 2002 as a longitudinal, pupil level database of characteristics and outcomes that the pupils gain in examinations. The data is largely collected from the Pupil Level Annual School Census (PLASC). For maintained schools in England, it is mandatory to submit a named pupil level return of the PLASC data. As noted in Jones and Elias (2006), since the dataset is a compulsory pupil level census, the NPD not only contains data on the individual pupils but also on their peers. Further to the characteristics of pupils, also contained in the NPD are data on pupils' outcomes in national assessments at ages 7, 11, 14 and 16. Further, school level data is contained, including the type of school, admissions policy of

[12] These GCSE tiers are examined in more detail in chapter 5.

the school, a grid reference for the school, number of full time teachers within the school, and whether there are pupils present who are boarders.[13]

Pupil level characteristics collected include the pupils' date of birth, age at the start of the academic year, their home postcode, their gender, ethnic group, first language, a measure of whether the pupil is a boarder or not, their exclusion status and a measure of low income with the free school meals (FSM) indicator. Free school meals are only available as follows:

> Children whose parents receive Income Support (IS); Income-based Job Seekers Allowance (IBJSA); income-related Employment and Support Allowance; support under Part VI of the Immigration and Asylum Act 1999; the Guarantee element of State Pension Credit; or Child Tax Credit, but who are not entitled to Working Tax Credit and whose annual income (as assessed by the Inland Revenue) that does not exceed £16,040; Working Tax Credit during the four-week period immediately after their employment finishes or after they start to work less than 16 hours per week are entitled to free school meals. Children who receive IS or IBJSA in their own right are also entitled to free school meals[14].

Pupils are assessed at 4 key stages through their school careers, at ages 7, 11, 14 and 16. The National Pupil Database (NPD) gives results of pupils in the key stage assessments. The structure of the data available when writing this thesis is shown in Figure 2, with results available for pupils who sat key stage 1 between 1998 and 2006[15], key stage 2 between 1996 and 2006, key stage 3 between 1998 and 2006 and GCSE between 2002 and 2006. The pupil-level data contained in PLASC, however, can only be linked to pupils who were in full time education when PLASC was initiated in 2002. Thus, the pupils who sat key stage 2 in 1996, for example, have no PLASC data.

For the releases of PLASC used in this thesis, with data up until 2006, it was not possible to observe any pupils through the entire assessment process in schools with the data in PLASC. The earliest exam data is available for key stage 2 in 1996, although this data is not linkable with pupil level demographic data from PLASC. Figure 2 shows the availability of data, and whether the data is linkable

---

[13] Jones and Elias (2006) give information on all variables available in the National Pupil Database.

[14] Current eligibility criteria for free school meals from http://www.direct.gov.uk/en/Parents/Schoolslearninganddevelopment/SchoolLife/DG_4016089 accessed 10/8/2009

[15] Chapter 4 uses an earlier release of the NPD with data only available until 2004.

with PLASC, or with future waves of the national pupil database. The longest that can currently be observed for one pupil is from key stage 2 to key stage 4, for students who were examined at key stage 4 from `02 to `06, and also from key stage 1 to key stage 3, for those students who were examined at key stage 3 in `05 and `06.

2.5 Selection of schools

In some LAs, a small proportion of maintained schools select pupils according to ability (e.g. grammar schools). This can cause problems for the analysis of data, as it polarises school populations within these selective schools, so pupils are only taught with similarly bright pupils. However, in these LAs, whilst the non-selective schools may have a comprehensive admissions policy, since the cream of the pupils has already been skimmed off, the ability levels of the pupils who are admitted will likely to be more strongly correlated with their peers than in a non-selective authority. Thus, in order to control for possible endogeneities caused by selection within LAs, it is not sufficient to simply remove selective schools from the sample. Rather than simply dropping selective schools, I define selective LAs as an LA in which at least 10 percent of its pupils in secondary schools are selected according to ability[16]. In order to control for the possible endogeneities, I omit these selective LAs.

In the empirical chapters of this thesis, I only consider schools which are classified to be community schools, foundation schools, voluntary aided, voluntary controlled and city technology colleges. I omit academies, as in the earliest release of the dataset, there are no results for pupils in academies. Also omitted are any special schools for pupils with special educational needs. In addition, any school that has records of having boarding pupils is dropped as well.

Table 1 shows the number of schools which have at least one pupil with a score at the relevant key stage at least once in the time period, broken down by key

[16] As defined in Atkinson et al (2006)

stage. This total includes all types of schools, including both special schools and independent schools. The second total of schools shows the number of schools which have at least 10 pupils in the cohort, which are city technology colleges, community schools, foundation schools, voluntary aided or voluntary controlled schools. By placing this restriction, 15 percent of the schools are dropped at key stage 1, 21 percent at key stage 2, and 40 percent at key stages 3 and 4. Further, I place the restriction that schools must be in a non-selective LA. By dropping these selective LAs, approximately 12 percent of schools are dropped at key stage 1 and key stage 2, whilst approximately 15 percent of schools are dropped at key stage 3 and key stage 4. Table 1 also examines the breakdown of these schools in non-selective LAs by type of school. It suggests that between 60 and 70 percent of schools are community schools, whilst between 15 and 25 percent of schools are voluntary aided. At key stages 1 and 2, approximately 15 percent of schools are voluntary controlled, and approximately 2 percent of schools are foundation, whilst at key stages 3 and 4, approximately 12 percent of schools are foundation, with approximately 3 percent voluntary controlled and 0.5 percent as city technology colleges.

2.6 Small schools

Chapters 4 and 6 use the NPD to examine the effect of a pupil's peer group on their outcomes. However, because the NPD only contains data of the school a pupil is taught in, and not the explicit classes that they are taught in, it is difficult to correctly identify the peer group that directly interacts with each pupil. In this section, I examine a property of infant and junior schools that allows me to identify the peers that a pupil is educated with in a subset of primary schools.

Infant schools cover the first three years of primary education, from age 4 to age 7 (i.e. in key stage 1). In infant schools since the start of the 2001/2002 academic year there has been a legal requirement that there should not be more than 30 pupils to a qualified teacher (the Education (Infant Class Sizes) Regulations 1998[17]). Effectively, this means that the maximum class size in

[17] See http://www.opsi.gov.uk/si/si1998/19981973.htm for more details. Accessed 10/8/2009

infant schools is 30. However, according to statistics from the Department for Children, Schools and Families[18] in 2009, 310 classes out of 53,160, or 0.6% of all key stage 1 classes, are taught in unlawful classes of 31 or more, whilst 580 classes (or 1.1% of all key stage 1 classes) were in lawfully allowed key stage 1 classes with more than 30 pupils. These lawfully allowed over-size classes are approved due to children being allowed entrance following appeals, being moved into a school's area after the start of the school-year or when the LA has placed a child with special needs into a school.

If infant schools follow this legislation, then we would expect to see a saw-tooth distribution of school cohort sizes. That is, we would expect to see peaks at sizes of 30, 60, 90 etc. This would indicate schools being filled until they reach the maximum size of 30 for a classroom, and if they have additional capacity, then starting a new class. Figure 3 shows the distribution of school-cohort sizes at key stage 1. It is clear that there is clustering at schools of size 30, 60 and at key stage 1, indicating that the schools are filling up the available spaces, and then stopping admissions.

Figure 4 examines the distribution of school sizes at key stage 1 before the introduction of the legal limit of 30 pupils to a class in 2002, and Figure 5 examines the data after the introduction. In Figure 4, there are markedly fewer schools that have 31 pupils in their cohort than have 30. Comparing this with Figure 5 there is a much more pronounced drop in the number of schools which have 31 pupils in their cohort compared with the number of schools which have 30 pupils in their cohort. Whilst the drop is less marked before 2002, at key stage 1, it seems a valid strategy to consider schools with 30 pupils as being schools which primarily teach all of their pupils within the school-year in one class.

At other levels, there is no legal maximum class size. However, Figure 6 shows the distribution of school sizes at key stage 2. There is a similar distribution to

[18] See http://www.dcsf.gov.uk/rsgateway/DB/SFR/s000843/SFR08_2009_ClassSizeCommentary.pdf for more details. Accessed 10/8/2009

that seen in key stage 1, but with a less pronounced falls after school sizes of 30 and 60. However, this evidence is sufficient to make the assumption that schools with 30 or fewer pupils consist of one class. Figure 7 shows the distribution of school sizes for key stage 3 and Figure 8 shows the distribution of school sizes for key stage 4. It is clear that in secondary schools, no such strategy is available to us, as the size of schools is much larger.

In order to consider classroom-level interactions, I use this characteristic of schools at key stage 1 and key stage 2 to examine schools which only include one class per cohort. That is, primary schools with 30 or fewer pupils present are likely to consist of only one class per academic year. As such, I define a small school as one that has 30 or fewer pupils in every observed cohort. Some schools within the dataset have unrealistically small cohorts, which may be the result of mixed age group classes within schools, which again introduces problems with measuring the peer group. In order to reduce this potential problem, I omit schools with fewer than 10 pupils the academic year.

Table 1 shows the number of schools that have 30 or fewer pupils in every observed cohort. At key stage 1, 4,300 schools are classified as small schools, with only one class per year, whilst at key stage 2, 3,534 schools are classified as small. As suggested by Figure 7 and Figure 8, there are only a very small number of schools at key stage 3 and GCSE (1 in each) which are classified as small schools, so in secondary schools it is much more difficult to accurately ascertain the peers with which a pupil is directly taught.

2.7 Summary statistics

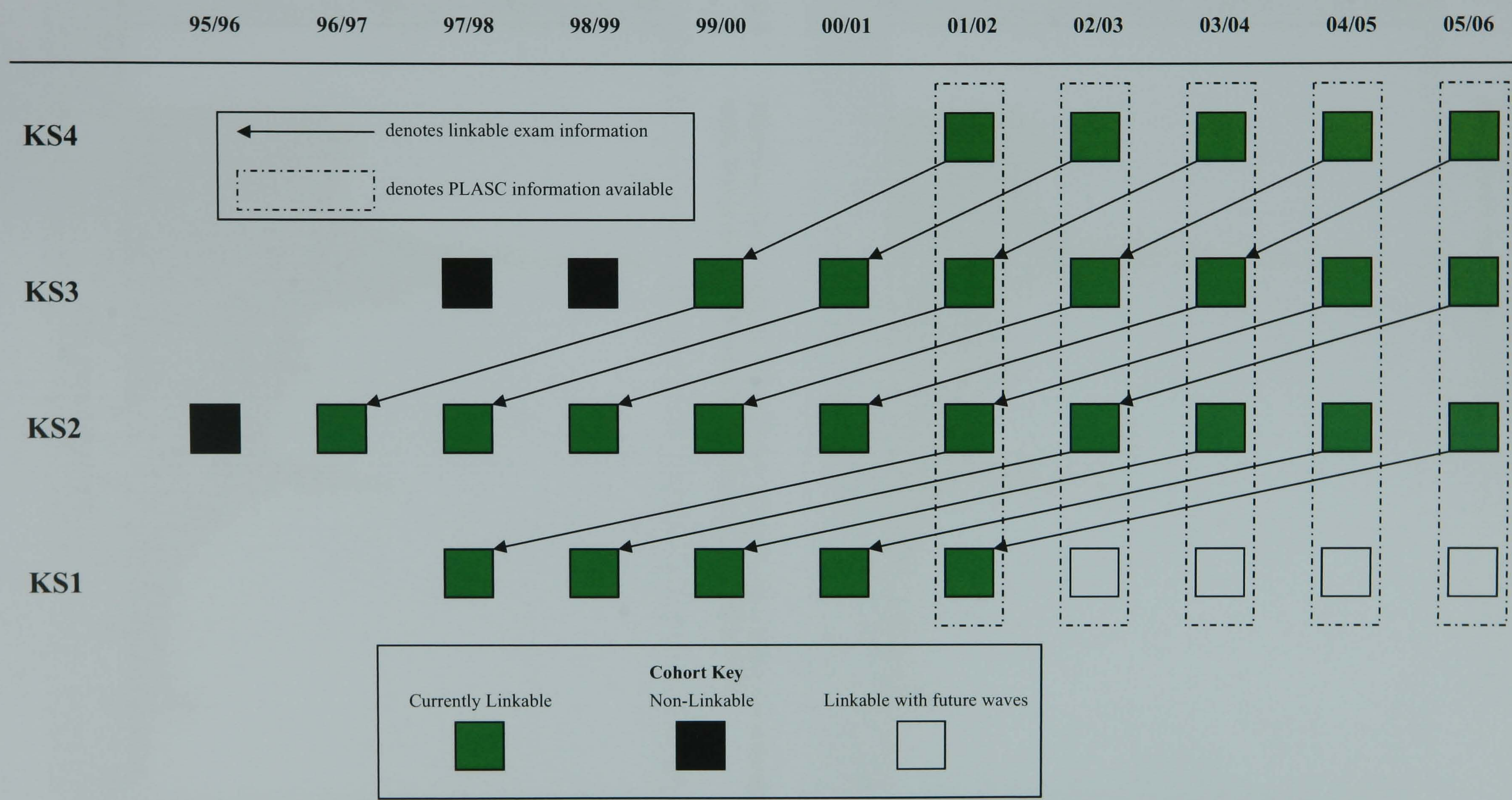
Table 2 shows selected summary statistics of demographics of pupils within the data, broken down by key stage. Key stage 1 has the smallest within school cohort size, with an average size of 38.2, at key stage 2, the average size is 42.6. Secondary schools are much larger, with the average cohort size at key stage 3 191.6 and at key stage 4 the average cohort size is 183.3. Looking at pupil level measures, between 50 and 51 percent of the pupils are male, with the majority of pupils also white British (approximately 83 percent). Approximately 2 percent of the pupils are Caribbean, 2 percent African, 2 percent Indian. 3 percent Pakistani, 1 percent Bangladeshi and the final 7 percent is made of other ethnicities. Between 14 and 19 percent of the pupils are eligible for free school meals, whilst between 2 and 3 percent of pupils have statements of special educational needs. Between 13 and 21 percent of pupils have special educational needs, but no statements. Finally, approximately 90 percent of the pupils have English as a first language.

Figure 1 The structure of school years in England

<table>
<tr><th>Age at Beginning of Academic Year</th><th>National Curriculum Year</th><th colspan="4">Type of School</th><th>Key Stage</th></tr>
<tr><td>4</td><td>Reception</td><td rowspan="3">Infant School</td><td rowspan="7">Primary School</td><td rowspan="4"></td><td rowspan="5"></td><td rowspan="3">Key Stage 1</td></tr>
<tr><td>5</td><td>1</td></tr>
<tr><td>6</td><td>2</td></tr>
<tr><td>7</td><td>3</td><td rowspan="4">Junior School</td><td rowspan="4">Key Stage 2</td></tr>
<tr><td>8</td><td>4</td><td rowspan="4">Middle School (Years 4-7)</td></tr>
<tr><td>9</td><td>5</td><td rowspan="4">Middle School (Years 5-8)</td></tr>
<tr><td>10</td><td>6</td></tr>
<tr><td>11</td><td>7</td><td rowspan="5">Secondary school</td><td rowspan="7">Secondary Schools</td><td rowspan="3">Key Stage 3</td></tr>
<tr><td>12</td><td>8</td><td rowspan="6"></td></tr>
<tr><td>13</td><td>9</td><td rowspan="5"></td></tr>
<tr><td>14</td><td>10</td><td rowspan="2">Key Stage 4/GCSE</td></tr>
<tr><td>15</td><td>11</td></tr>
<tr><td>16</td><td>12</td><td rowspan="2">Sixth Form</td><td rowspan="2">Key Stage 5</td></tr>
<tr><td>17</td><td>13</td></tr>
</table>

Notes: Primary school consists of a combined infant and junior school. Some secondary schools contain sixth forms; others simply cover up to the end of key stage 4. Middle schools are not present in all local authorities. There are two main types of middle schools, from years 4-7 and from years 5-8. This thesis does not consider education beyond the end of compulsory education at the end of key stage 4.

Figure 2 The structure of the cohorts available in the NPD and PLASC[19]



[19]Adapted from http://www.bris.ac.uk/depts/CMPO/PLUG/userguide/cohorts.pdf accessed 10/8/2009

Figure 3 **Distribution of school sizes at key stage 1, focussing on schools with fewer than 100 pupils, covering years 1998-2006**

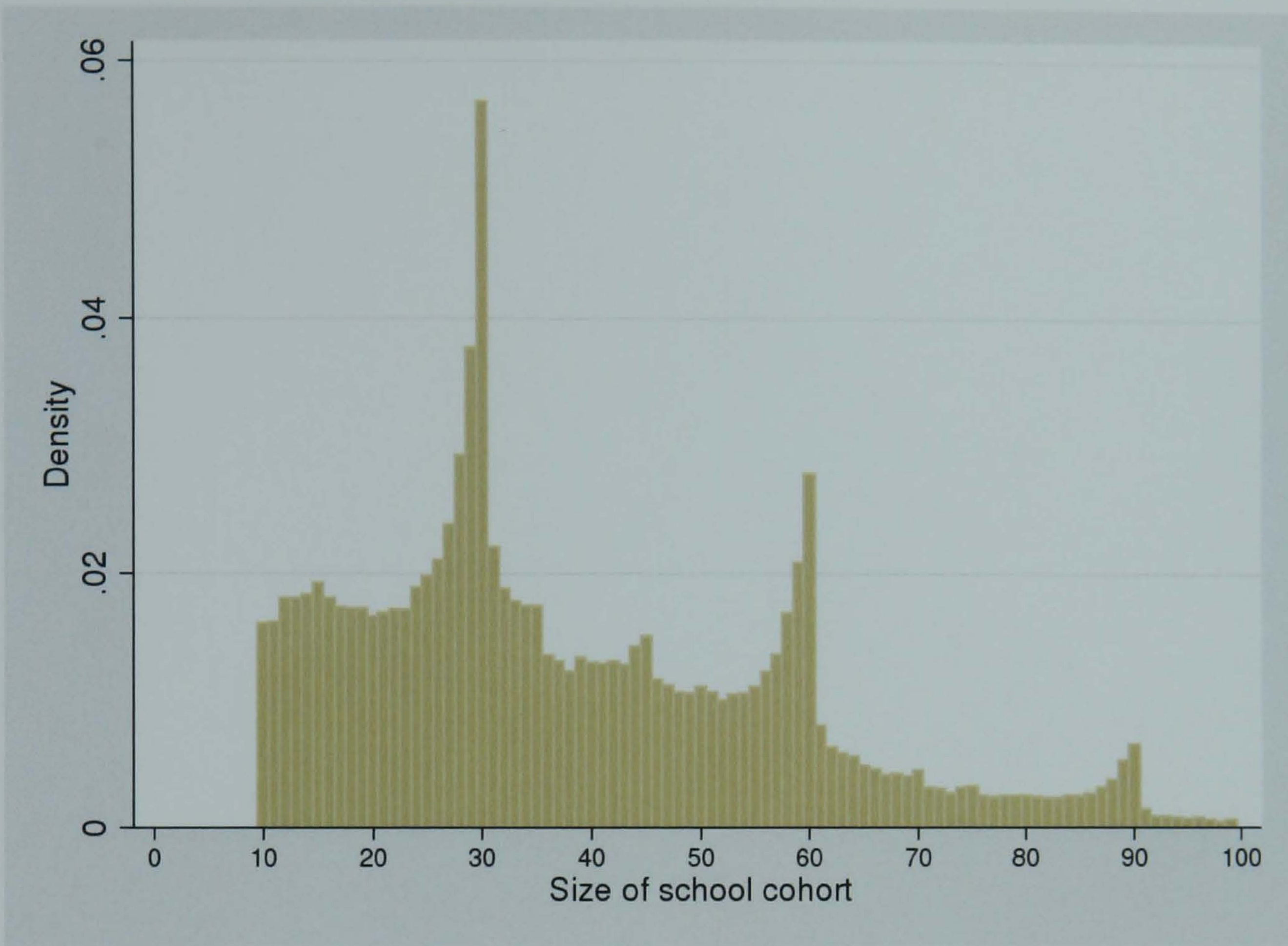


Figure 4 **Distribution of school sizes at key stage 1, focussing on schools with fewer than 100 pupils, covering years 1998-2001**

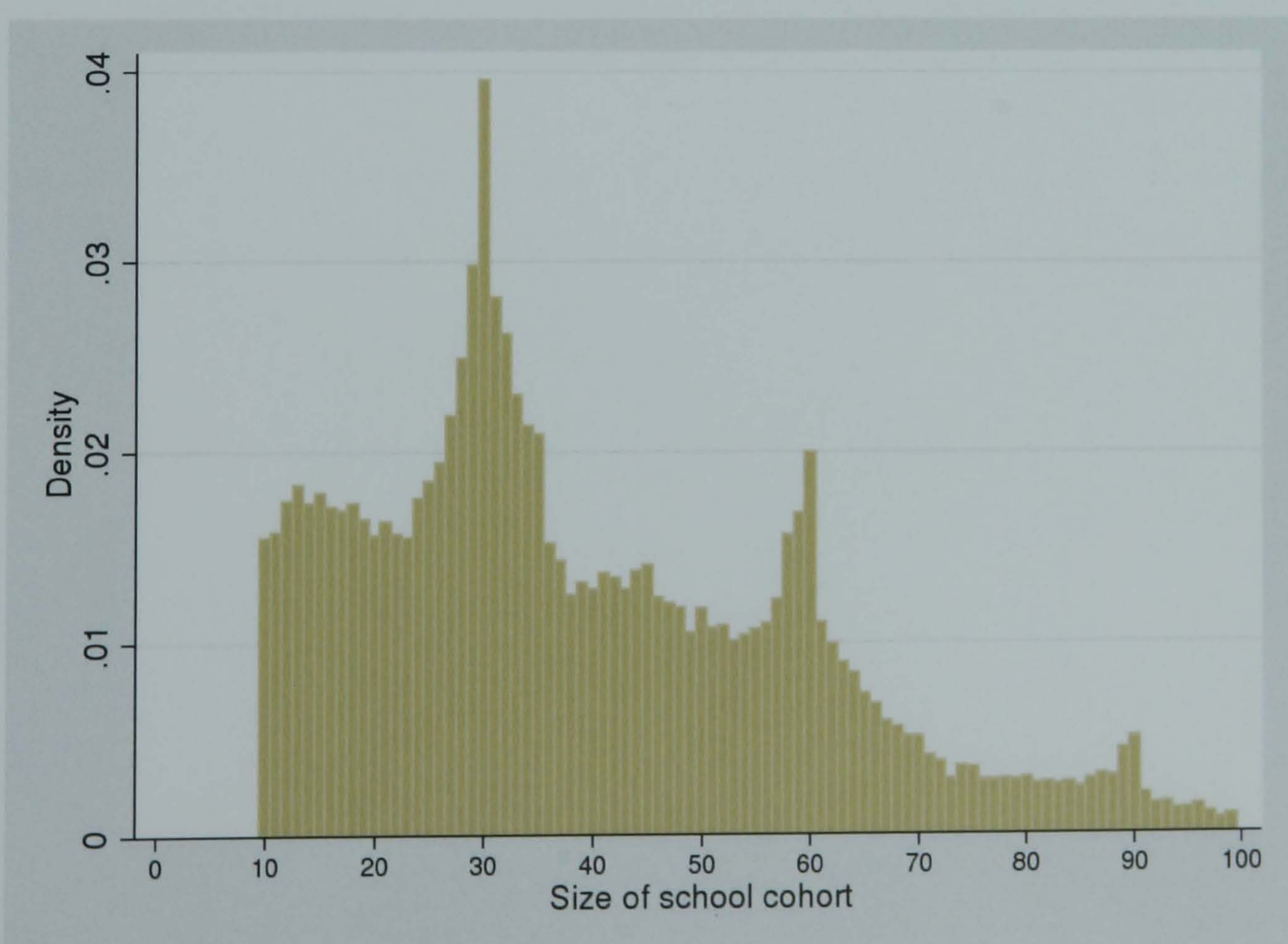


Figure 5 **Distribution of school sizes at key stage 1, focussing on schools with fewer than 100 pupils, covering years 2002-2006**

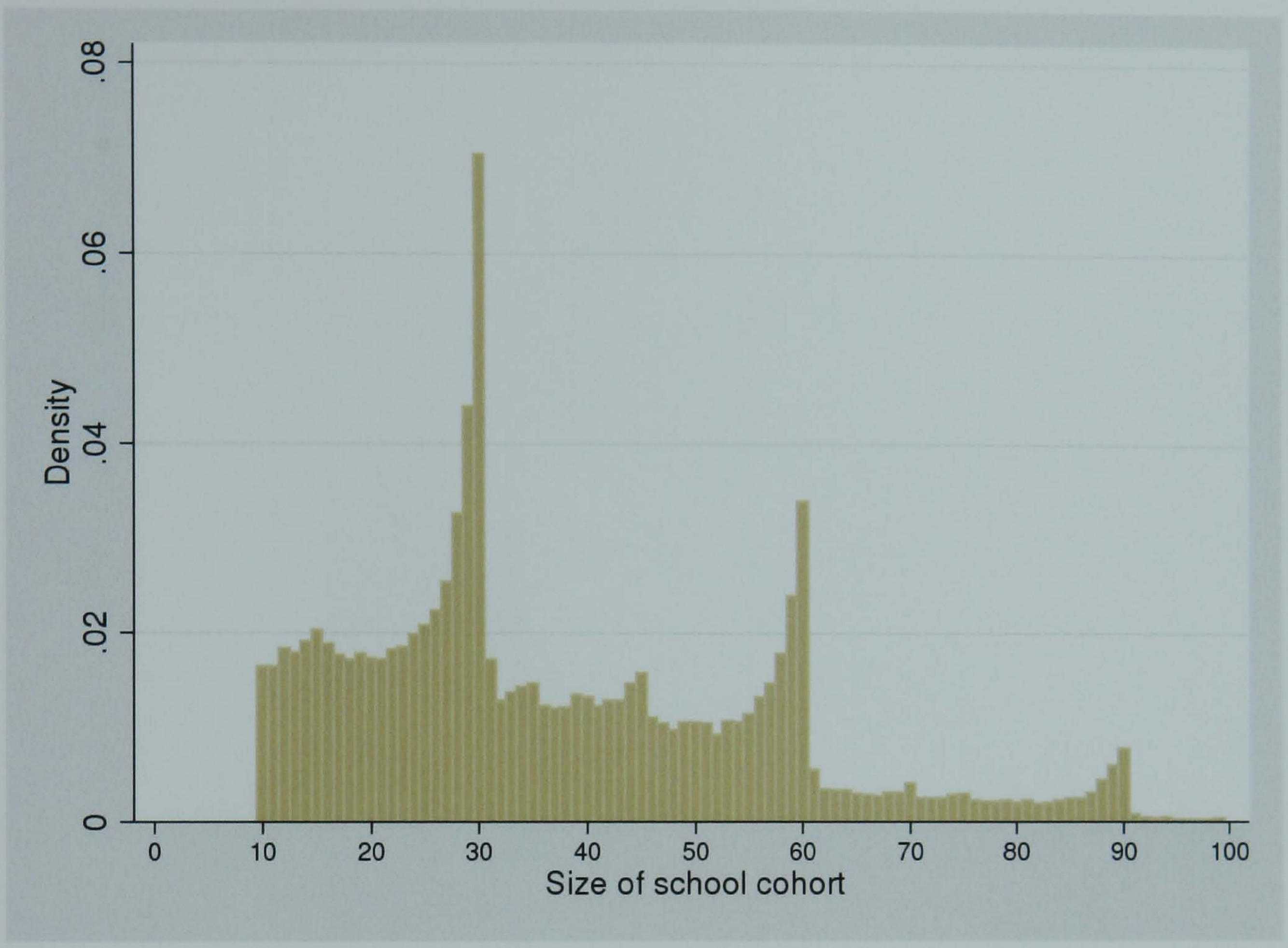


Figure 6 **Distribution of school sizes at key stage 2 focussing on schools with fewer than 100 pupils, covering years 1997-2006**

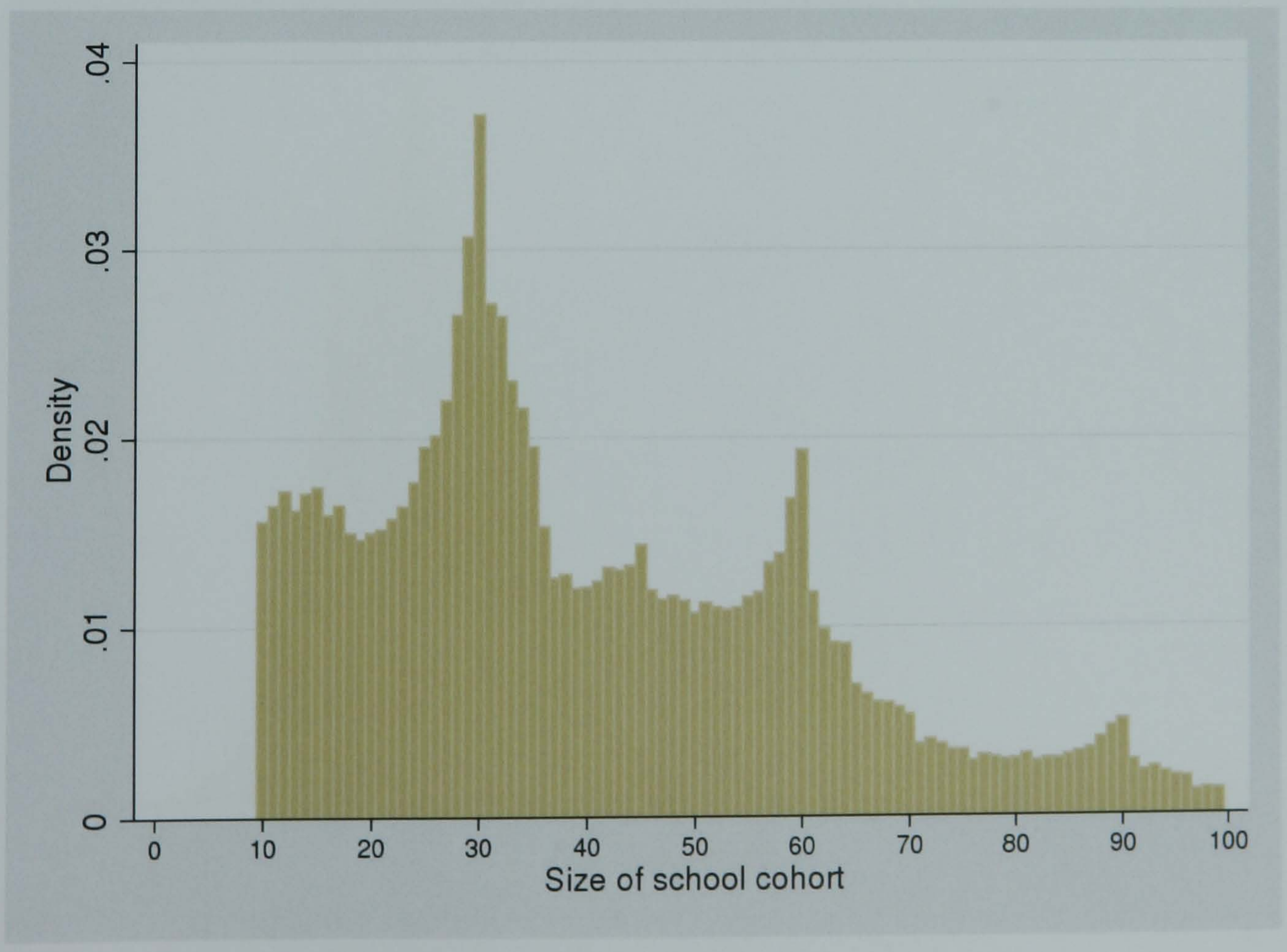


Figure 7 **Distribution of school sizes at key stage 3, covering years 1998-2006**

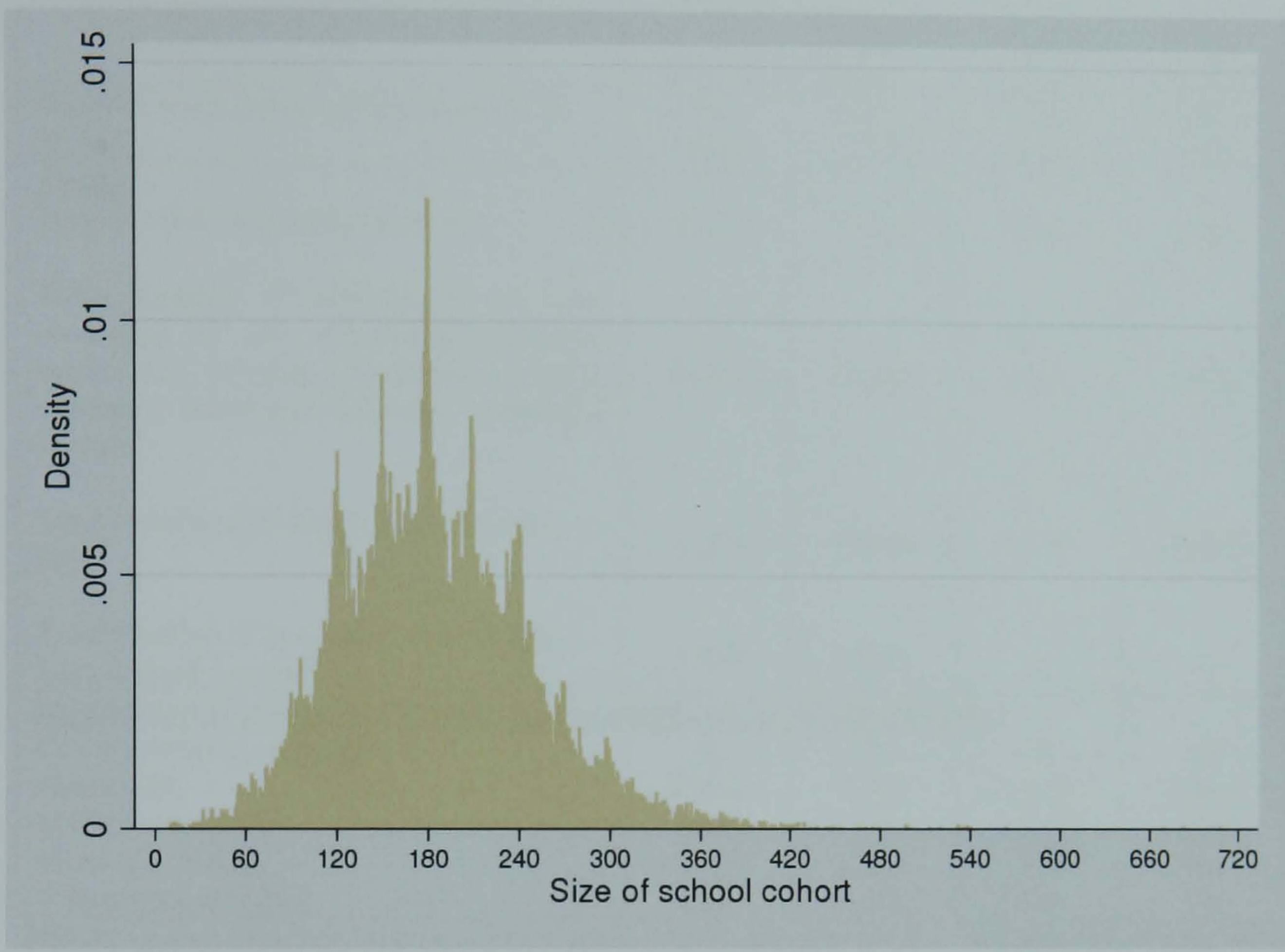


Figure 8 **Distribution of school sizes at key stage 4 covering years 2002-2006**

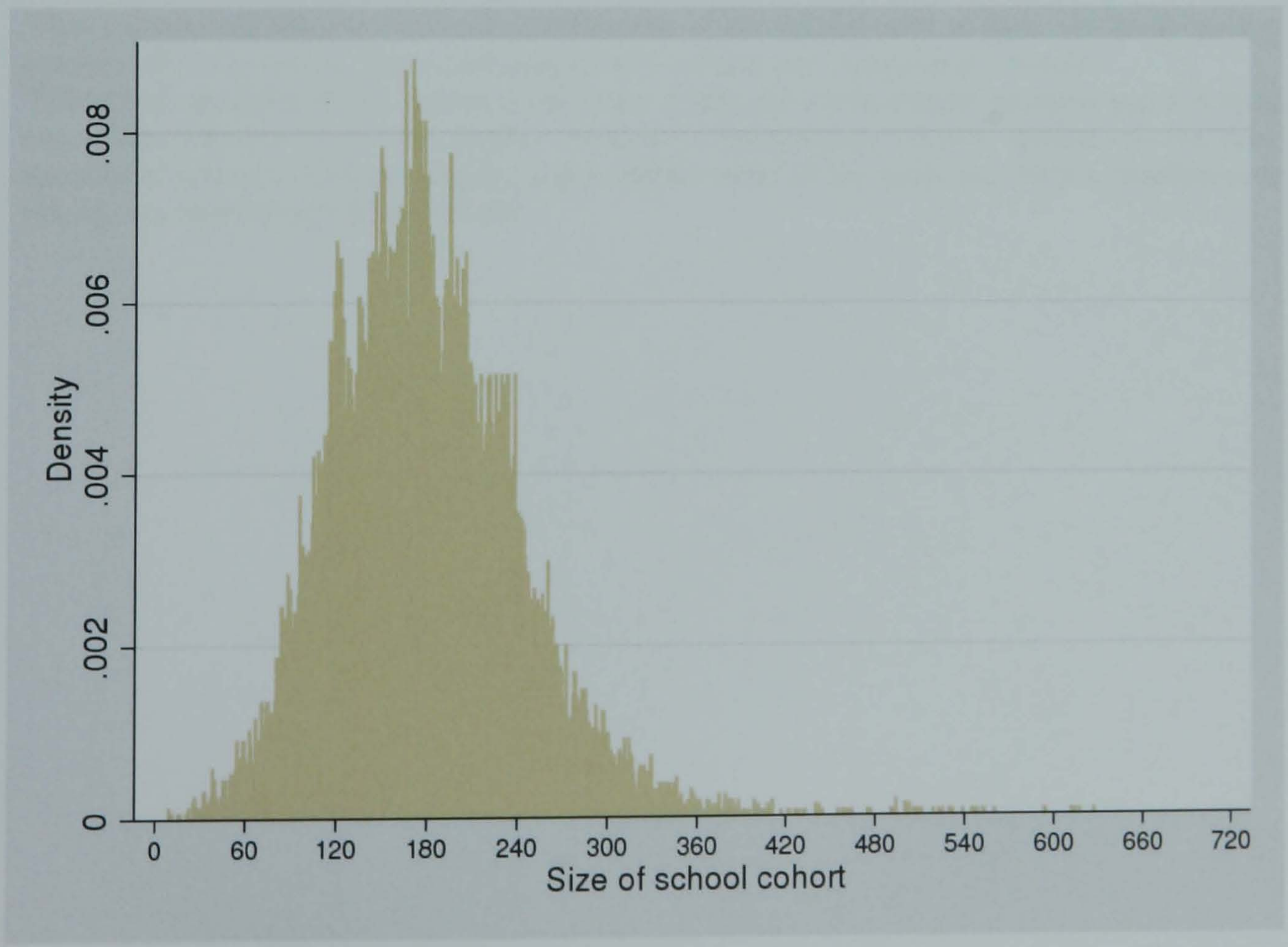


Table 1 Totals of schools

	Key Stage 1	Key Stage 2	Key Stage 3	Key Stage 4
Years of exam data available linkable to PLASC	1998-2006	1997-2006	2000-2006	2002-2006
Totals				
Total number of schools observed[a]	18,855	19,217	5,305	5,234
Total Number of schools in all LAs, consisting of city technology colleges, community schools, foundation schools, voluntary aided and voluntary controlled schools[b]	16,124	15,169	3,181	3,142
Total Number of Schools in non-selective LAs[b]	14,276	13,386	2,730	2,695
Total Number of Schools in non-selective LAs defined as small[b]	4,300	3,534	1	1
Breakdown of number of schools in non-selective LAs by school type				
City Technology colleges	0	0	13	14
Community	8,812	8,200	1,839	1,812
Foundation	208	237	336	333
Voluntary aided	3,184	3,079	470	460
Voluntary controlled	2,072	1,870	72	76

Notes: Unit of observation is a school-cohort within each key stage. The number of schools is the total number of schools, for which there are up to 9 observations at key stage 1, 10 observations at key stage 2, 7 observations at key stage 3 and 5 observations at key stage 4. A small school is defined as one that has 30 or fewer pupils in each observed cohort.

[a]This includes all schools observed at least once in the dataset with at least one pupil having results at the relevant key stage, including special schools and independent schools.
[b]Schools are included if they appear at least once within the national pupil database and if at least one cohort contains 10 or more pupils. Academies, independent schools, special schools, non-maintained schools, offshore schools, pupil referral units, secure units and Welsh establishment schools are omitted here from the data.

Table 2 Summary of selected key indicators in the data

	Key Stage 1	Key Stage 2	Key Stage 3	Key Stage 4
School level measure				
Average size of cohort	38.7	42.6	191.6	183.3
	(20.5)	(25.3)	(61.3)	(61.7)
Pupil level measures				
Proportion of pupils eligible for free school meals	0.186	0.171	0.158	0.136
	(0.389)	(0.377)	(0.365)	(0.343)
Special Educational Needs				
Proportion of pupils with a statement of special educational needs (SEN)	0.018	0.027	0.026	0.023
	(0.133)	(0.162)	(0.159)	(0.150)
Proportion of pupils with SEN but without a statement	0.207	0.187	0.157	0.128
	(0.405)	(0.390)	(0.364)	(0.335)
Proportion of pupils who are male	0.511	0.508	0.506	0.502
	(0.500)	(0.500)	(0.500)	(0.500)
Proportion of pupils with English as a first language.	0.892	0.907	0.909	0.907
	(0.310)	(0.290)	(0.287)	(0.291)
Proportion of pupils of each ethnicity				
White British	0.829	0.839	0.835	0.825
	(0.377)	(0.367)	(0.371)	(0.380)
Caribbean	0.015	0.016	0.015	0.016
	(0.123)	(0.124)	(0.123)	(0.125)
African	0.019	0.015	0.015	0.016
	(0.136)	(0.123)	(0.122)	(0.124)
Indian	0.023	0.023	0.024	0.025
	(0.149)	(0.151)	(0.152)	(0.156)
Pakistani	0.031	0.025	0.025	0.025
	(0.172)	(0.158)	(0.156)	(0.156)
Bangladeshi	0.013	0.011	0.010	0.010
	(0.113)	(0.103)	(0.101)	(0.100)
Other	0.070	0.070	0.076	0.084
	(0.256)	(0.256)	(0.265)	(0.277)

Notes. Ethnicities shown are only those that contribute more than 1% of the total pupil population. Other ethnicity includes: other black background, other Asian background, Chinese, white and black Caribbean, white and black Asian, any other mixed background, any other white background, Irish, Gypsy/Romany, traveller of Irish origin, any other ethnicity and information missing or not obtained. The average size of the cohort is the average across school-cohorts of the number of pupils within the cohort. All other statistics are estimated based on pupil level data. Figures given are means with standard deviations in parentheses.

Chapter 3 Peer Effects and Literature Review

3.1 Introduction

In this thesis, I am interested in the effect of a child's peer group on their academic outcomes. An obvious place to start with this analysis would be to consider an educational production function, such as the one suggested by Todd and Wolpin (2003). Todd and Wolpin (2003) propose a cumulative production function for children's academic outcomes as follows:

$$A_{tij} = A_t[F_{ij}, S_{ij}, \mu_{ij}, \varepsilon_{ij}] \quad (1)$$

where attainment for child i in school j is a function of the prior family inputs, F, their school inputs S, underlying ability μ and an error term ε. In order to relate this production function to the effect of a child's peer group on their outcomes, it is important to consider one of the most influential educational reports of the 20th century, the Coleman report (Coleman et al (1966)) which discussed aspects of the educational production function, and concluded that the largest individual factor in children's outcomes was determined by their family inputs and background. However, considering the school level factors, they suggested that the school level inputs at time t could be decomposed into a cumulative of teacher inputs (T), inputs from the child's classroom based peer group (C) and other school inputs, such as facilities and ethos of the school (S).

$$S_{ijt} = T_{ijt} + C_{ijt} + S_t \quad (2)$$

Their results suggested that "Attributes of other students account for far more variation in the achievement of minority children than do any attributes of school facilities and slightly more than do attributes of staff." (Coleman et al (1966, 86). That is, of these school level factors, the make up of the peer group has the largest effect, followed closely by the characteristics of the teachers, and outweighing the effect of other school factors. These peer factors could include students' socio-economic status, the ability of the peer group, the mix of ethnicities within the peer group, or other such attributes. Whilst the order of magnitude of these effects have been questioned by recent research, such as Kramarz et al (2009), it is still important to assess how a child's peer group will influence their academic and social outcomes.

These peer effects could, however, be transmitted through a number of different mechanisms. For instance, Lavy and Schlosser (2007) examine the effect of a more female peer group on academic outcomes, and they suggest that a change in the gender make-up of the peer group also has an effect on the behaviour of children within the classroom, with more boys leading to a more disruptive environment which is not conducive to learning. That is, the presence of more girls within the classroom is correlated with the behaviour of the class as a whole, which is in itself correlated with the learning experience for the pupils. Conversely, since there is a large body of evidence of girls outperforming boys in English tests[20], it is possible that teachers will focus their teaching towards the majority of the class. If the majority of the class are male, then the teachers are likely to focus towards the male students' level, whilst if the majority of the class are female, the teachers are likely to focus towards the female students' level. Similarly, changes in the ability make-up of the peer group can influence how the teacher teaches the class. If there are a high proportion of low achievers within the class, the teacher will inevitably have to go more slowly, as is appropriate for the majority of the class. However, with a high proportion of high achievers, teachers would be able to work much faster, but if this mechanism is in place, it would imply that it would be very easy for students who are a long way from the ability of their peer group to fall behind. Another possible mechanism to explain the effect of the ability, or prior attainment of the peer group, is an ability spillover. That is, being in a class with a more able pupil simply leads to you doing better due to the presence of this higher ability pupil. A final possibility, as with the gender make-up of the classroom, is that a pupil in a class with peers who are exerting high effort (i.e. pupils who achieve higher results from exerting higher effort) is likely to benefit as this can create a better learning environment, and the influence of the peer group could put pressure on other pupils to exert similar high effort.

In this thesis, I consider the effect of two different aspects of the peer group on children's academic outcomes. In chapter 4, I examine the effect of the gender make-up of the peer group on outcomes, whilst in chapters 5 and 6 I examine

[20] See, for instance, Machin and McNally (2005)

different methods of estimating the effect of a more able peer group on outcomes at ages 11 and 16 within the classroom and within the school. In this chapter, I discuss the literature on measuring the effects of a child's peers on their outcomes. I begin by discussing the early literature on the effect of a child's peer group on their outcomes, and then discuss problems inherent with estimating these effects, as identified by Manski (1993). I then move on to strategies that are used to deal with these problems: random assignment of pupils, instrumental variables techniques and finally studies which control for heterogeneity of pupils. I conclude by offering a summary of the literature.

3.2 Peer effects

Beyond the Coleman report, early studies examined the effect of differences in the racial make-up of the peer group. These include Winkler (1975) who found differential effects of the composition of peer groups on different races, and Summers and Wolfe (1977) who found that both black and non-black pupils benefited from a more balanced mix of black and non-black pupils, and also commented that students who tested at or below grade level were helped by being in a school with high achievers.

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.

However, Manski (1993) identifies problems with trying to measure the effect of peer groups on outcomes. Peer effects can be split into three different types; endogenous effects, correlated effects and exogenous effects. In the case of endogenous effects, decisions made by the individuals within the peer group directly affect the decisions made by other members of the peer group. The second effect is a correlated effect, which is largely due to members of a peer group having some trait in common, which in turn influences the outcomes of the peer group due to the individuals making similar decisions due to their similar characteristics. The final type of effect is an exogenous effect, where one's actions depend on the exogenous characteristics of one's peers. Learning outcomes appear to be endogenous effects, as for instance, a child's desire to work hard could affect other children in the class's decision whether to work hard or to misbehave and vice versa. Manski (1993) discusses the problems inherent with such endogenous peer effects when trying to infer the effects that members of a reference group have on its own members (the reflection problem), and argues that it is not possible to draw inferences on effects unless one has prior knowledge of the make-up of the reference group. Furthermore, he argues that studies using apparent random distribution may experience bias to the apparent peer effect if there are unseen family characteristics that are in common with the reference group. As such, it is difficult to accurately gauge the effect of a child's peer group on their own outcomes.

3.3 Strategies for dealing with possible endogeneity of peer groups.

In order to confront the problem of endogeneity in the peer group measure, there are three main strategies employed to remove this problem; random assignment, instrumental variables techniques, and finally, attempting to control for the large amounts of correlated family characteristics, by also controlling for pupil level and family level heterogeneity.

In this section, I examine the literature utilising these three strategies for identifying the effect of a peer group measure, firstly examining random assignment, then instrumental variables techniques and finally moving on to studies which control for heterogeneity.

3.3.1 Random assignment

Several studies use apparent random assignment to get around the problem of selection. Whilst I am particularly interested in the school age studies, there are some interesting methodologies offered in order to deal with possible endogeneity for studies using college aged students. Concentrating first on school-age studies; Hoxby (2000) uses multiple strategies for assessing the effects of race and gender. The initial strategy makes use of her argument that there is a credibly exogenous variation in the distribution of female pupils across cohorts within a grade, and simply uses the raw proportion of girls as a measure of the peer group. That is, year on year changes in the proportion of girls in a school cohort are essentially random, and not affected by neighbourhood or school characteristics. By utilising this strategy she finds that if all of the peer effects operate through peer ability, then a 1 point increase in peer ability should lead to a rise in pupils' scores by between 0.3 and 0.5 points in reading, and in mathematics a raise of between 1.7 and 6.8 points. However, she goes on to argue that this is far too large an effect to be credible, and so other mechanisms must be in operation. When she looks at the effect of race, whilst each year does draw from the same population, one could expect more non-random shocks in distributions than for gender due to parents moving into a school catchment area. I discuss the Hoxby's (2000) instrumental variables strategy in the next section. Lavy and Schlosser (2007) consider a similar gender strategy to Hoxby (2000), but their estimates indicate that, as with Hoxby's estimates, the effect of a change in the gender make-up of the peer group is not caused by a change in the ability of the peer group, but rather due to the change in behaviour of the peer group. Kramarz et al (2009) also consider a similar strategy to Hoxby (2000), and try to estimate the effect of various peer groups to calculate the contribution of various inputs to a pupil's test scores. They consider that year on year changes within the school-grade will essentially be exogenous. Their results suggest that the effect of the peer group is significant, but provides a smaller contribution than school level contributions.

Gibbons and Telhaj (2008) examine English data in order to estimate the effect of a more able peer group. Since there is no system of random assignment to schools in England, they take advantage in year on year changes in the formation of a peer group at age 11. These year on year changes are similar in essence to those examined by Hoxby (2000). They find that whilst linear estimates indicate little effect of a more able peer group, these estimates are misleading, as their results indicate that whilst higher and middle ability students gain from a more able peer group, those at the bottom of the distribution are actually disadvantaged.

Ammermueller and Pischke (2009) use similar credibly random methodology by examining primary schools in Germany, France, Iceland, the Netherlands, Norway and Sweden. They suggest that within schools, allocation to classes is credibly random, and so within school estimates should give unbiased estimates of the effect of a more able peer group. Whilst their initial within school estimates are smaller than the OLS estimates, they also show that controlling for measurement error increases the estimate of the effect of the peer group within schools to the same size as their original OLS estimates. Their results suggest a 1 standard deviation increase in the ability in reading of the peer group is associated with a 0.17 standard deviation increase in the reading ability of individual fourth graders.

Vigdor and Nechyba (2007) use data on 5th grade students in North Carolina. Their basic OLS specifications find significant and persistent effects of a more able peer group, with the ability of the peer group in the 5th grade still affecting outcomes of pupils in the 8th grade, with these results robust to the inclusion of school fixed effects. However, they suggest that the results are not causal, due to selection bias. In order to try to get around the selection bias, they consider schools that have credibly exogenous shocks to the peer group composition, caused by the opening of new schools, or other shifts in the boundaries for attendance. They further suggest that a change in the make-up of the peer group within 'feeder schools' has no significant effect on the test-score outcomes, which seems contradictory to the initial findings of significant positive effects of a more able peer group.

Hoxby and Weingarth (2005) take advantage of the re-allocation of children to different schools in Wake County in the USA in order to estimate the effect of a more able peer group on outcomes. Wake County moved up to five percent of the school population each year, initially to balance racial make-up of the schools but post 2000, the movements were to balance income composition. This created numerous natural experiments, with pupils being exposed to new peers. Their results suggest that provided achievement of the peer group is properly controlled for, then racial and income characteristics only have either small or insignificant effects on outcomes. However, their results also suggest that there is a monotonic, but non-linear effect of a more able peer group on outcomes, with the presence of more able peers providing benefits to students.

Kang (2007) takes advantage of quasi-randomization and IV methods for pupils in South Korea, where the law states that pupils entering middle school should be assigned randomly by a lottery within their school-district. Kang also argues that ability grouping within schools in Korea is rare, and so the peers that children experience are credibly random, and their peer group is essentially exogenous. In order to show causality, Kang (2007) utilises an IV strategy by using the peer achievement in science as an instrument for their mathematics achievement and vice versa. The results imply that a 1 standard deviation increase in the mean mathematics score of the peer group leads to a 0.3 standard deviation increase in an individual's mathematics score, and also go on to suggest, in contrast to Hanushek et al (2003) that weak students would benefit from being educated with a wide range of abilities, whilst high ability students benefit from being grouped with similarly able students, but this finding is contradicted by Betts and Shkolnik (2000) who suggest that no ability group benefits from being grouped with similarly able students.

Lavy et al (2008) attempt to estimate the effect of a large proportion of pupils who are repeaters rather than being educated with their original cohort. In order to estimate this effect, and to avoid problems with endogeneity, they utilise cross cohort differences within schools. They argue that an increase in the proportion of pupils who are repeaters leads to a deterioration in the quality of teaching practices, as well as an increase in violence and disruption within classrooms, as well as deterioration of inter student and student-teacher relationships. Furthermore, they show that the effect of repeaters is most damaging to students with low socio-economic status. However, they also argue that this is not strictly an endogenous peer effect due to spillover from low ability to regular students, but rather a contextual effect of having more repeaters within the classroom. A similar strategy is employed 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.

Imberman et al (2009) take advantage of the forced evacuation over one million people affected by the impact of Hurricanes Katrina and Rita in 2005. Many children were forced to relocate across the South East of the United States, and were swiftly enrolled in new schools, with 20,000 pupils enrolling in schools in Houston whilst Louisiana schools enrolled 196,000 new pupils. Therefore, there was an exogenous influx of pupils into schools, which allowed the estimation of the effect of the evacuee children on the non-evacuee children. Their models suggest that an increase in the ability of the peer group is associated with an increase in individuals' outcomes, and similarly a decrease in the ability of the peer group reduces individuals' outcomes. However, their analysis also suggests that the effects are non-linear when assessing the effect of mean ability. Furthermore, they show that if the evacuees were undisciplined, this also had a detrimental effect on the behaviour of the indigenous children, which supports the model of the "bad apple" affecting its neighbours.

Considering now college level studies, but it is necessary to remember that the mechanisms involved with the peer effects in these studies may be different from the mechanisms involved with school age peer effects. Sacerdote (2001) uses the fact that first-year students are randomly assigned room-mates and dorm-mates to argue an exogenous peer group of the room-mate and further to the entire dorm, finding effects on grade point averages at room-level and decisions on joining social groups at both the room and dorm level, but does not find effect of peers on academic decisions.

Zimmerman (2003) uses assignments of first year students to room-mates in Williams College to create students` peer groups. Students at Williams College are assigned according to a housing preference form asking questions on smoking, attitudes to noise etc. He finds little evidence of a link between the students' housing questionnaire and their Grade Point Average once previous SAT results are controlled for, and the only questions that make a significant difference are those that have the least weighting on housing decisions. Winston and Zimmerman (2003) use a similar strategy for three schools. For two of these the three schools find very similar results. Both of these studies conclude that there is little or no credible effect on pupils at the top of the SAT distribution, but that pupils in the middle of the SAT distribution suffer a negative effect if they are housed with students in the bottom 15% of the SAT distribution.

Stinebrickner and Stinebrickner (2006) also use the apparent random assignment of first year students to room-mates. However, they also introduce the proviso that, according to evidence from the Berea Panel Survey, only 37% of the students list their dorm-mate as one of their four best friends, but their room-mates are the people that they spend most time with, whether more through necessity than choice. Thus, they argue that previous studies which have simply used direct effects may have been looking in the wrong places. They use the time-use survey to try to identify where the possible peer effects could come from room-mates and whilst direct teaching and discussions of subjects may occur, the median student only spends 20 minutes a day discussing topics and since they may not be studying the same subjects, this activity may in fact be costly. However, Stinebrickner and Stinebrickner (2006) argue that time spent studying may act as a behavioural effect. They find that the high school grade point average acts as better proxies for time spent studying than the American College Test. They find effects of room-mates' high school grade point average and family income on individual students' outcomes.

Hoel et al (2006) use a very similar strategy using random assignment to classes for the compulsory humanities 110 unit, and apparent random assignment of room-mates. However, they do admit that there may be a bias associated with this assignment as they cannot assess possible non-random groupings of students due to the students' housing preference forms. With this possible bias, they find strong effects of ability on outcomes with room mates and dorm-mates, but do not find evidence of effects at a classroom level. Parker et al (2008) expands the analysis performed in Hoel et al (2006) to examine student enrolment into core courses. Again, they find little evidence of high ability peers on other students' outcomes outside of the core course. They suggest that the mechanisms through which peer effects occur are not explicitly through ability, but are instead based on attitude and personality.

Carrell et al (2009) also use random assignment to peer groups, utilising random assignment to peer groups of about 30 in the United States Air Force Academy (USAFA), where the students are required to spend most of their time interacting with their assigned peer group, including living in adjacent dorms, eating studying and competing in intramural sports together, whilst they only have limited opportunities to interact with the other peer groups. They find larger estimates of the effect of a more able peer group on outcomes than those seen in other studies, equal to a 0.45 point increase in grade point average, on a 4.0 scale, attributable to an increase in the peer group average SAT verbal score by 100 points. They furthermore show that the largest effects are seen in mathematics and science, with no significant effect seen in foreign languages or physical exercise. However, they do recognise that since there are behaviours that are encouraged in the USAFA, such as trying to foster teamwork, that the peer effects observed may be larger than those seen in other institutions. Furthermore, since negative social activities, such as underage drinking, carry strict punishments, it is possible that potentially negative peer influences are eliminated.

3.3.2 Instrumental variables

Whilst Hoxby (2000) initially utilises credibly exogenous changes in the makeup of the gender and race makeup of the peer group, she argues that the racial makeup may not be random due to preferences of living in an area with a similar racial make-up. Hoxby's solution is to use the cohort to cohort changes in unexpected shocks across years by using the residual from a regression of the proportion of the cohort that is a certain race against a constant and a time trend as an instrument for the change in peer groups. A second approach taken by Hoxby is to regress the average achievement for each group within cohort (e.g.

black males in the third grade) against the proportion of each group in the cohort and obtain the residuals, and if there are no peer effects, then a regression of the residual against the median score for the group interacted with the group proportion should yield coefficients of 0. Hoxby finds that both males and females tend to perform better in reading when they are in more female classes, and as an example of the results suggests that an all female class would score about one fifth of a standard deviation higher in reading.

Robertson and Symons (2003) use an educational production function using the past peer group, schooling and family inputs experienced by the child along with previous levels of achievement for UK data. The paper uses a new set of instruments, regions of birth, to estimate their production function. They consider effects of classmates from different socioeconomic groups and the effects of being in "streamed classes", which entails pupils being taught in sets with similar ability pupils based on outcomes in previous assessments. They show that, apart from the very lowest ability pupils in mathematics classes, streaming helps the higher ability students and worsens outcomes for the lower ability pupils. Furthermore, they discuss the advantages of having peers of higher socioeconomic status.

Angrist and Lang (2004) utilise a combination of strategies, and examine the effects of relocating (mostly) black students from inner-city Boston to schools in the more affluent suburban areas using the (voluntary) Metco bussing programme. In order to identify the effects of the change in the peer group, they use an instrumental variable based on the predicted number of Metco students in the class. Whilst the black students have a significantly lower baseline ability level, they find little or no effect to the non-Metco pupils, although there is a suggestion that there is a small negative effect on black girls in the host district, but these are very much short term effects.

Lefgren (2004b) uses an instrument constructed from the R-squared value obtained from within school regressions of students' prior ability on dummies for what class they are taught in. This R-squared value is interacted with the pupils' prior ability to form the instrument. Lefgren finds significant, but small effects of peers' average ability on pupil outcomes which are considerably smaller than the OLS estimates. I discuss Lefgren's methodology in more detail in Chapter 5.

Maurin et al (2005) take advantage of the correlation between outcomes and age within year to estimate the effect of the peer group on outcomes. They use the percentage of pupils born in each month to try to identify the effect of the peer group. Their analysis suggests that the peer effect is non-linear, but they cannot disentangle whether the effect they are observing is from being with peers of higher ability, or whether the observed effect is from being grouped with older children.

Clark (2005) uses a novel instrumental variable technique using a dummy indicating whether you are in a school with elder pupils (a middle school) rather than concluding primary education with younger peers in a primary school as an instrument. This study finds statistically significant peer effects in terms of anti-social behaviour such as drinking, smoking or drug abuse.

Dills (2005) uses an OLS and IV strategy to estimate effects of introducing a magnet school into a school district which "creams off" the highest ability students. As an instrument, they use a dummy for whether there is a magnet school in the district. It argues that the creaming off leaves plausibly exogenous variation in the quality of classmates remaining in the `regular' schools. It finds that the loss of the high ability pupils does have a negative effect on the performance of the lower achieving pupils left in the regular schools, thus implying the existence of some positive peer effects.

3.3.3 Controlling for heterogeneity

Other studies try to control heterogeneity by including a large number of pupil-level variables. Zimmer and Toma (2000) try to remove unobserved family heterogeneity by including parents' occupation, parents' education, gender, presence of siblings and language spoken at home. However, as with all such studies this study does run the risk that there are other heterogeneities which are not observed, which could cause a bias. They do find effects that are larger on low ability than high ability which are robust across countries but not school type, but one must bear in mind that the bias from the endogeneity of the peer measure may still be present. Similarly, Ding and Lehrer (2007) try to control for large amounts of heterogeneity by using a rich Chinese dataset to estimate the effect of a more able peer group. They find robust and significant evidence of a benefit to students from an increase in the ability of the peer group, although their results suggest that the effects are not homogenous. Their results suggest that pupils with the highest test scores gaining the most benefit from a more able peer group.

Henderson et al (1978), whilst not explicitly saying this, also try to control for heterogeneity, by including school, teacher and family characteristics in their regressions. Their overall results show that a heterogeneous peer group helps overall average student performance, although at the expense of the higher achieving students. They do introduce the question of whether high achieving students should make sacrifices to their outcomes in order to help society as a whole. They also claim that the variance of the peer group does not affect the students' outcomes. Hanushek et al (2003) try to utilise this same technique, but include individual and school by grade fixed effects as well as family and school characteristics in order to remove the major components of between student and between family that do not vary over time such as underlying ability, parental attitudes, material inputs etc. They hope that this technique removes sources of bias from either mis-measured or omitted family and individual factors. They find significant effects of peer achievement for all students, but quantile regressions reveal that pupils in the top quartile may be less responsive. However, they find little evidence of any effects of peer income or variance of peer ability, although they do admit that the free school meals measure of income is a very noisy one. They also contend that it is possible that since schools may have a fixed catchment area, and thus a fixed community to draw its pupils from, the inclusion of school fixed effects may lower the variance of the peer group. Also, they examine other possible mechanisms for peer effects by examining the effects of differences in class variance of scores, but find no systematic effects.

Burke and Sass (2008) 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.

McEwan (2003) also employs school level fixed effects on schools in Chile. He argues that whilst in Chile there may be far less academic tracking than in the US, it is still possible that within schools pupils are set according to ability and other unseen factors which would create a bias in the possible peer effects. He goes on to ascertain that the most important peer factor was the classroom average of the mother's education, whilst average father's education has a smaller, but still significant effect on outcomes.

Betts and Zau (2004) control for endogenous changes in students' recent achievement and utilise a large number of pupil, school and classroom characteristics including student fixed effects. They show that year to year changes in the peer group have a significant effect on students' rate of learning. They analyse these effects at both classroom and grade level, and find significant effects for English and mathematics at the Classroom level and find less significant effects for English at the Grade level. Furthermore, they investigate the symmetry of peer effects around zero, and find that an increase in peer group from the mean is likely to gain less than an identically sized decrease in peer group score.

Schneeweis and Winter-Ebmer (2007), rather than using the achievement level of pupils as the peer measure, use the peer-groups' socioeconomic composition as a proxy for attitudes and learning related activities. The problems of omitted variables are countered by including a number of powerful explanatory variables affecting academic achievement and peer group formation, and also, a school-level fixed effects model is utilised. This may suffer from self-selection bias, as one would expect similar socioeconomic parents to choose similar schools, although this will be controlled for by the school level fixed effects, but there also may be some other unseen heterogeneity, such as parents' teaching attitudes that influences the child's setting. Austrian schools are very heavily stratified, so within a school with a certain socioeconomic class, there would be expected differences such as this parental teaching attitude which certain types of parents might be willing to use. However, they find evidence that there are substantial positive effects of the peer groups' socioeconomic composition, whilst quantile regressions tend to imply that the effects are strongest for the lowest ability students, but have little effects on the highest achievers. They thus suggest that it is best to lower stratification or streaming within schools to create a more heterogeneous make up within schools. However, this is somewhat contradicted by Robertson and Symons' (2003) analysis of UK data, who imply that high ability students perform better in streamed classes with other high ability students, although they both conclude that lower ability students lose out by being in a very stratified class. Slavin (1987) uses a collection of analyses to perform meta-analysis, finding achievement effects that were effectively zero and no real effect of ability grouping using a comparison of means for homogeneous and heterogeneous sets. However, this approach is criticised by Hallinan (1990) as contradicting the wealth of literature suggesting positive effects of ability grouping for high achievers and negative effects for low achievers. Hallinan argues that comparing means is effectively meaningless as in high ability sets, teaching would be geared to quicker learning and conversely low ability sets, teaching would be geared to slower learning, whilst mixed teaching would be geared to the middle of the set.

3.4 Summary

Whilst there is some disagreement between papers, the general consensus of the literature is that an increase in the ability of a child's peer group will lead to an improvement in the outcomes for the child. Authors suggest a range of the effect for a more able peer group from little or no effect (Angrist and Lang (2004), Lefgren (2004b)), to moderate effects (Hoxby (2000), Hanushek et al (2003), Ammermueller and Pischke (2009)) to much larger effects (Kang (2007)). The sizes of these effects vary according to the country observed and the methodology employed to try to remove any endogenous element of the peer group. Each of these methodologies inevitably have both positive and negative aspects when trying to measure the effect of the peer group. For random assignment of students, if the students are truly randomly assigned, then we do not have the problem of correlated attributes of the students causing bias to the estimates. However, there is a strong possibility in these studies that the allocation is not truly random (e.g. for college room allocation, some aspects of students are taken into account when allocating the rooms), which would negate the apparent randomness of the allocation. Also, this methodology would only be able to offer predictions of the effect of a peer group in similarly allocated random group, so it would be more difficult to generalise to the entire population, due to the complexity of mechanisms involved with peer effects. The IV strategy offers a good way of assessing the effect of a peer group. Provided the instrument set is valid, then IV should give unbiased estimates of the effect of the peer group on pupils' outcomes. However, it is difficult to find instruments that are truly exogenous, and which are correlated sufficiently with the endogenous variable. If the instruments are invalid, then bias is no better than for OLS, and the estimates are less efficient. Finally, considering the studies which control for large amounts of heterogeneity; these studies again have the opportunity to provide unbiased estimates of the effect of the peer group, but only if the very strong assumption that we have no unobserved heterogeneity which is correlated with both the peer group and individual outcomes. As such, these estimates always need to be treated with caution.

Chapter 4 Girl Power?

An analysis of peer effects using exogenous changes in the gender make- up of the peer group

4.1 Introduction

The Coleman Report (1966) identified that a child's peer group as an important determinant in their own outcomes. Hoxby (2000) suggests that in US schools, a more female peer group leads to improved outcomes for both boys and girls within schools, a finding that is re-iterated in Lavy and Schlosser (2007).

This chapter builds on work by Hoxby (2000)[21] using exogenous changes in the gender make-up of the within school peer group to estimate the effect of a child's peers on their educational outcomes. Hoxby's initial strategy utilizes the credibly exogenous variation in the distribution of females across cohorts within a grade, using the raw proportion of girls as a measure of the peer group. She then combines this with the test-score gap between girls and boys to estimate the effect of an exogenous change in the ability of the peer group. This chapter utilizes the same strategy, but builds upon it in three important ways. First, I take advantage of the fact that in England, there is a legal upper limit of 30 for class sizes for children in infant schools[22], as discussed in chapter 2. I use this fact to separate schools that appear to only have one class per cohort to estimate classroom level effects rather than school-level effects. Secondly, I investigate whether there is any bias to the estimates by including a measure of the average socioeconomic status of the male and female pupils separately using the proportion of boys (or girls) who receive free school meals (FSM) within the cohort in the school. Finally, I examine the effect of a more female peer group on the average progress from one national assessment to the next.

This analysis uses data on English pupils from the Pupil Level Annual School Census (PLASC) and the National Pupil Database (NPD). This data includes pupils' results from national assessments and demographics of the pupil, such as age within year, gender, ethnicity and free school meals status. These

[21] Also published in abridged form as Hoxby (2002)

[22] Infant schools cover ages 4 to 7.

assessments are key stage 1 sat at age 7, key stage 2 sat at age 11, key stage 3, sat at age 14 and GCSE sat at age 16[23].

I find significant negative effects of a more female peer group for males in English at all levels of assessment, and significant positive effects of a more female peer group on both boys and girls in mathematics and science, although these positive effects largely disappear post age 11. The omission of the socioeconomic status in the initial models has no significant bias on the coefficient on the proportion of the school-cohort that is female. The value added model shows strong significant positive effects of a more female peer group between ages 7 and 11 in English for both girls and boys, and between ages 11 and 14 for girls in mathematics and science. Furthermore, considering the effect of more females in the class as a proxy for changes in ability, I demonstrate that the magnitudes of the effects are too large, and of the wrong sign, to be explained by small changes in ability.

I first consider the literature specific to the effect of the gender make-up of children's peer groups, and then move on to examine the methodology and the data. I then examine the results and offer some discussion.

4.2 Literature

In chapter 3, I discussed the inherent problems associated with estimating peer effects, due to endogeneity and correlations of neighbourhood characteristics. In this section, I discuss specific literature relating to the gender make-up of the peer group on outcomes.

Much of the previous literature has tried to explicitly consider the effect of single-sex classrooms on outcomes. For example Marsh and Rowe (1996) find little effect of single sex classes, with male pupils feeling less favourable to single sex classes. In the UK Malacova (2007) employs multilevel methodology, and finds an advantage for girls educated in single sex classrooms, but with this

[23] A more detailed description of the English schools system is given in chapter 2

advantage decreasing according to prior ability, whilst the advantage of single sex education decreased for boys according to the level of school selectiveness.[24]

As discussed in chapter 3, Hoxby (2000) and Lavy and Schlosser (2007) examine the effect of the gender make-up of the peer group within schools and classrooms, and both find significant positive effects of a more female peer group in English and mathematics. These findings are partially backed up by Whitmore (2005), who finds positive effects of a high proportion of girls in kindergarten through to the second grade on the outcomes of both boys and girls. However, Whitmore (2005) also suggests that in the third grade there is evidence that boys do worse in a class with a high fraction of girls. Hansen et al (2006) find that female dominant and equally mixed groups perform better than male dominated groups.

Hoxby (2000) suggests a possible mechanism for her observed effects that differential ability of boys and girls in the classroom could be accountable for these effects. However, the results she finds are too large to be solely explained by changes in the ability of the peer group. Furthermore, Lavy and Schlosser (2007) show that their results move in the opposite direction to what would be expected due to peer effects caused by a more able peer group.

In order to validate these previous studies, I investigate whether the size and direction of effect in this study is credible to be attributable to differences in the ability of the peer group caused by the differential achievement of girls and boys. In order to estimate the effect of a more able peer group using changes in the gender make-up of the class, it is necessary for there to be a significant difference in outcomes for girls and boys in school outcomes.

Hallinan and Sorenson (1987) consider reasons for the differential achievement levels in mathematics, with boys holding the advantage. Whilst they conclude that mathematics teaching within stratified groups does not have a differential effect for girls and boys, they do find that the initial grouping decision is indeed

[24] A more complete analysis of the advantages and disadvantages of single sex education is provided in Campbell and Sanders (2002)

influenced by the gender of the pupil. Male high achievers are far more likely to be assigned to a high achieving group than female high achievers, indicating some unseen factors also affecting the grouping decision (or alternatively just some prejudice against girls in mathematics).

Shibley Hyde et al (1990) carry out a meta-analysis of research on the magnitude of the gender gap in mathematics using 100 separate studies. Whilst they found a male dominance in the subject, this was decreasing over time. Also, some bias was present due to a self-selection problem. When considering performances based on samples of the entire population, females in fact had an advantage albeit a negligible one. However, as samples became more selective, a gender gap became apparent favouring males, which has reduced over time. This implies that either males are more likely to be the ones to gain the higher grades, or that boys are more likely to drop mathematics if they are not good at it.

Machin and MacNally (2005) examine the education system in England, and show that girls outperform boys in all schools from primary school onwards, particularly in English. In mathematics, the story is a little more complicated, with little difference between the proportion of boys and girls reaching the target grade at age 11. However, they show a clear advantage for girls in English. Similarly, Burgess et al (2004) find that the gender gap is largely seen with girls outperforming boys in English, with very little difference in performance in mathematics and science between equivalent male and female students.

Gorard et al (2001), examining the gender gap in Wales, suggest that all pupils enter education on an equal footing, which is supported by little difference in outcomes between boys and girls at key stage 1. They also suggest that there is little gap between boys and girls in mathematics at key stages 2 and 3, but until recently, there had been a gap in favour of boys at GCSE. For science, they suggest there is little gap between key stages 1 and 3, with a historical gap between girls and boys at GCSE. In English, again, they find little difference at key stage 1, but they find at older levels, there is a much larger proportion of girls achieve grade D or higher.

The majority of the literature suggests there is a significant difference in outcomes for boys and girls in English, suggesting a change in the gender make-up of the classroom would be associated with a change in the average prior attainment of that class. There is also some suggestion of a gender gap in mathematics, but it has moved from an advantage for boys to an insignificant difference in recent years.

4.3 Methodology

In this chapter, I use the same basic methodology as used by Hoxby (2000), utilising idiosyncratic changes in the proportion of pupils in the school cohort that is female as a measure of the peer group. This can then be combined with the difference in outcomes associated with the gender of the pupils to try to estimate the effects of a more able peer group on outcomes, and to investigate whether there are more mechanisms in play than simply higher ability peers helping to increase the performance of the rest of the peer group.

Beginning with the educational production function, as set out in chapter 3, an individual child's academic outcomes are a function of their family inputs, school inputs, underlying ability and an error term. The gender make-up of the peer group is a school level input, and in order to identify the effect of the gender make-up, I begin with an individual-level educational production function. The model uses the assumption that any school j at a given key stage, g, has an average outcome for male (female) pupils, which is constant across cohorts, c, and differences from this mean can be explained by peer group effects, other factors not correlated with the peer effects and some unobserved random factor. So, for a female pupil i, there is a production function thus

$$A_{i,gjc} = \mu_{female,gj} + \gamma_{female} p_{gjc} + \alpha_{female} X_{i,gjc} + \varepsilon_{i,gjc} \qquad (3)$$

where μ is a school-level fixed effect, consisting of a constant, an average school outcome and a school-level fixed effect, p is the proportion of pupils in the school-cohort that is female, which is the peer group influence that we are interested in, and X represents other pupil-level exogenous and constant variables, which also includes year and key-stage dummies. The dependent variable A is the individual's score within the school year. The levels represented are i for each individual pupil, *female* (or *male*), g representing grade, or exam-level being taken, j representing the school attended and c representing the cohort the pupil is a member of. This production function assumes that male and female students experience different effects from the proportion of pupils who are female, as well as other exogenous factors. I later try to control for changes in the demographics by including a measure of relative deprivation in the school; that is the proportion of pupils who receive free school meals (FSM). There may be a possibility that female pupils with a low socioeconomic status have a different effect to females with a high socioeconomic status, so to try and control for this effect, I enter the proportion of male pupils receiving FSM and the proportion of female pupils receiving FSM separately.

The exogenous and constant variables, X, consist of fixed family background effects (F), the pupil's underlying ability (U) and various exogenous factors (χ), including year dummies and dummies for the level of the examination.

$$X_{i,gjc} = F_{i,gjc}a + U_{i,gjc}b + \chi_{i,gjc}c + e_{i,gjc} \qquad (4)$$

Since the identification strategy operates at a school level, when taking means, I assume that F and U are drawn from a population with unchanging demographics. Furthermore, I assume that these effects are uncorrelated with the probability of a child being female, and any time-invariant effects should not bias the effects of a more female peer group.

This individual model (3) can be collapsed down to a school level average. However, since males and females have different average outcomes, whilst a school average would be directly affected by the proportion of pupils in the school that is female, I use separate specifications for male and female pupils, which will not be affected in this way.

$$A_{female,gjc} = \mu_{female,gj} + \gamma_{female} p_{gjc} + \alpha_{female} X_{female,gjc} + \varepsilon_{female,gjc} \qquad (5)$$

The motivation behind this model is that at a given exam, a school has an average outcome that is achieved, and each year there is a variation around this mean, that is influenced by the proportion of pupils that is female and other exogenous effects.

In order to remove the school-level fixed effects, I take first differences across cohorts within a given key stage,

$$A_{female,gjc} - A_{female,gj(c-1)} = \gamma \Delta p_{female,gjc} + \alpha \Delta X_{female,gjc} + \varepsilon_{female,gjc} - \varepsilon_{female,gj(c-1)} \qquad (6)$$

$$\Rightarrow \Delta A_{female,gjc} = \gamma \Delta p_{female,gjc} + \alpha \Delta X_{female,gjc} + \Delta \varepsilon_{female,gjc} \qquad (7)$$

This identification strategy depends on there being no endogenous component of the change in gender make-up of a school. Since the distribution of genders of pupils can be seen as credibly random, then it can be argued that changes in gender makeup should also be credibly random, and as the size of school increases the proportion of girls should tend to the national average.

There is a potential problem with this strategy. Since there is no data on classroom level interactions within the school, it is possible that the magnitude of effect could be mis-estimated. That is, a pupil who attends a school with a large proportion of pupils who are female may not experience this grouping within the classroom. In order to address this possibility, I use the fact that in England there has been a legal limit placed on the size of infant class sizes (ages 4 to 7) of 30, which was instituted in 2002, as discussed in chapter 2. This allows me to examine schools with 30 or fewer pupils within the school-year as a proxy for schools that teach their pupils in one class per year. I show later that this can be extended for infant schools for the period before 2002. Whilst there is no such limit imposed on junior schools (serving pupils aged 7 to 11), many junior schools are linked to an infant school, and follow a similar policy with regards classroom allocation. As shown in chapter 2, there is a similar structure of school sizes in junior and infant schools. Thus, I define a small school to be one that has thirty or fewer pupils in every observed cohort, whilst a large school is defined to be one that has more than thirty pupils in every observed cohort. Pupils over the age of 11 are educated in much larger schools, and so we cannot extend the strategy further.

Thus far, I have simply considered using the levels that students receive from examinations at ages 7, 11, 14 and 16. These levels are highly correlated with other, unobserved factors such as family background and neighbourhood affluence. In order to try to control for this, I also examine a value added[25] score within subject. The value added measure I use is simply the test score achieved by an individual pupil at one key stage subtracted from the score obtained at the subsequent key stage. For instance, the value added at age 11 is simply the test score at age 7 subtracted from the test score at age 11[26]. In order to examine the effects of a more female peer group, I would like pupils to remain in the treatment group for the whole period between examinations. Due to the structure of schools in England almost all pupils (98.6%) have changed schools between key stage 2 and key stage 3, whilst few pupils change between key stage 3 and key stage 4 (3.1%). Wilson (2004) shows that there is a low correlation between test scores and value added, and thus the effect of school level inputs may be better viewed using this value added score.

In order to try and keep the treatment group constant across the treatment period, I consider only the pupils who stay in the same school between key stage 1 and 2 and between key stages 3 and 4. However, the vast majority of children in England change schools between year 6 (key stage 2) and year 7 (key stage 3), and so without any further information about the school attended, I make the assumption that the pupils are at a fixed school in years 7 to 9, which will be the case for the vast majority of pupils. Thus, for the key stage 2 to 3 measures, I consider those pupils who have moved schools between the exams. The number of pupils who appear in the sample, and the number omitted are shown in Table 3.

---

25 This is not the same as the value added score used for nationally published league tables.
26 There are other methods of calculating value added.

Since I am not interested in the time or grade effects in the structural model. I simply include year and grade dummies in the first difference equation.

4.3.1 Tests of robustness

It is possible that particular schools have policies on admission that makes the proportion of pupils that is female an endogenous measure, or that variation in the gender makeup of the school follows a non-random pattern due to some other external factor. In order to examine this possibility, I use a similar strategy to Hoxby (2000). That is, for every school within grade, I perform a regression of the proportion of pupils that is female against a linear time trend and a constant and obtain the R-squared value. The order of the years within the schools is then randomised, and a further regression is performed, again on a linear, but randomised, time trend and the R-squared value is obtained. The R-squared values from the two regressions are compared. Schools with a ratio of greater than 1.20 for the real time trend R-squared value to the false time-trend R-squared value are dropped from the sample. Whilst Hoxby (2000) also included non-linear trends, since I only have 3 time observations for GCSE, this is not possible at this level in my data, due to a lack of degrees of freedom. This results in approximately half of the schools being dropped, and a comparison of the results for the sub-sample and the full sample is reported in Table 11, Table 12 and Table 13.

Finally, in order to ensure that the linear model of the peer effects is the correct specification, I use a regression including the interaction between the change in the proportion of pupils that is female and the quartile that this is in.

$$\Delta A_{female,gjc} = \gamma\Delta p_{female,gjc} + \delta_1\Delta p_{female,gjc}q_1 + \delta_2\Delta p_{female,gjc}q_2 + \delta_3\Delta p_{female,gjc}q_3 + \Delta\varepsilon_{female,gjc} \qquad (8)$$

I then use an F-test to test that $\delta_1 = \delta_2 = \delta_3 = 0$

4.3.2 Weighting of data

This analysis uses several specifications, with some consisting of results from several key stages. This raises two issues. First, since the dependent variable is created by taking a mean of pupils' test scores, simply using this score unweighted would lead to a mis-specification of the model, as large schools would necessarily have the same weight in the model as small schools. Thus, the first part of the weighting is the number of pupils used to create this average score. The second issue is raised when I pool multiple key stages in the analysis, as for instance, there are only 3 observations of GCSE results, whilst there are 8 years of key stage 2 results. Since I take first differences, there is one fewer observation in the OLS specification, and so, I consider the number of cohorts less one. Thus, the second part of the weighting is to divide the weights by the number of cohorts, less one, that are observed for each key stage assessment. Furthermore, this only gives the weight required for each individual year, rather than for the change between years, so in order to deal with this, I take the average of the weightings for consecutive years.

i.e. The weight is calculated thus:

$$W_{male,gjc} = \frac{(N_{male,gjc} + N_{male,gjc-1})/2}{C_g} \qquad (9)$$

Where N is the number of male (female) pupils in the school and C is the number of cohorts observed at level g.

4.4 Data

This chapter uses an early release of the NPD, as discussed in chapter 2, with pupils' examination results up until 2004, giving 7 years of data at key stage 1, 8 years at key stage 2, 5 years at key stage 3 and 3 years at key stage 4. Schools are selected as per the criteria set out in chapter 2. Beyond this criteria, it is apparent that some of the schools appear to have vastly different numbers of pupils from one year to the next. Since this chapter considers year-on-year changes in the make-up of the school, it is important that schools with incredibly large changes in their size should not bias the results. In order to prevent these outlying schools from adversely affecting the results, I only consider schools that lie within the 1st to the 99th percentiles of cross cohort changes in school sizes. That is, schools which have an improbably large change in size from one year to the next are removed from the sample. In real terms, at key stage 1, I only consider schools that fall by a maximum of 20 pupils from one year to the next and rises by 18, at key stage 2, a maximum fall and rise of 21, at key stage 3, a maximum fall of 43 or a rise of 52 and at key stage 4, a maximum fall of 34 and a rise of 54.

Further, some schools have very large (or very small) proportions of girls in the school. In order to remove the possibility that some of these schools have some sort of endogenous selection policy based on gender, schools that lie outside of the 1st to 99th percentile of the gender mix (after single sex schools are dropped) are also dropped. This leads to a range of the proportion of pupils that is female between 16.66% and 80%.

Finally, in order to have a consistent sample across the time series, only schools that have observations for every time period are included. Thus, any school that closed, (or opened or failed to report results) during the time-period of the data is omitted. This leads to a total of 10,180 schools at key stage 1, 9,031 schools at key stage 2, 2,083 schools at key stage 3 and 2,227 schools at key stage 4.

Science at key stage 4 needs to be treated carefully. Not all pupils are assessed in the same way for science. There are three possible structures that are examined for science; one single award, covering all of physics, chemistry and biology, a dual award, which gives the students two identical grades, or up to three separate sciences. Thus, a student may receive 1, 2 or 3 grades at key stage 4 science. As such, to create a comparison across pupils, I consider the mean of their science scores.

The raw data is presented in terms of national curriculum levels achieved by the pupils in the specific key stage, which should be comparable across years. In order to make the results easily comparable across key stages, the raw, pupil-level, results are standardised by subject and level to a mean zero and standard deviation of one.

4.5 Summary statistics

Table 4 shows summary statistics for the entire sample, for secondary schools and for each individual key stage, whilst Table 5 shows summary statistics for small primary schools.

The scores from English, mathematics and science key stage assessments are presented in a weighted form, as described above. The proportion of girls within the cohort and the size of schools are weighted slightly differently, with the number of cohorts observed at each level used as the weighting. Whilst this does not affect the statistics within key stages, when they are pooled it does place more weight on the larger secondary schools. Science appears to have a lower sample size in the pooled specification simply because science is not assessed at key stage 1, whilst English and mathematics are.

Since all of the individual key stage results are based on means of normalised results centred at 0 with standard deviation 1, it is possible to compare the mean scores between key stages. Looking at the pooled data, it can be seen that on average girls perform much better than boys in English, but in mathematics and science, there is little or no difference, with boys initially holding an advantage, although girls overtake them by the time they reach key stage 4.

At key stage 1, girls have a significant advantage in English, whilst there is little difference between the genders in mathematics. At key stage 2, there is still a significant advantage for girls in English, whilst in mathematics and science, the boys hold a small advantage. At key stage 3, the gap between the genders is increased in English, and boys still hold a very slight advantage in mathematics and science. However, this changes slightly at key stage 4, with girls maintaining a large significant advantage in English, but taking a small lead in mathematics and science as well.

The gender mix in the schools remains constant at approximately 48% to 49% female throughout, with cohort sizes within school of approximately 40 at key stages 1 and 2 and approximately 180 at key stages 3 and 4. This may make inferences at a school level much harder at the secondary level due to the fact that whilst there may be a larger proportion of female pupils in one school than another, individual pupils may not feel the effect of this due to a lack of within school interaction. That is, at a cohort level there could be a large proportion of girls, but this may not permeate down to the classroom level, whether due to ability setting or some other mechanism.

Table 5 shows the summary statistics for key stage 1 and key stage 2 in small primary schools. As with Table 4, the proportion of the cohort that is female is approximately 0.49. By placing the restriction on considering single classroom primary schools, I keep 23 percent of the schools.

4.6 Results

In looking at the results, I start by looking at a specification that includes the full sample of schools, and all of the available key stages, followed by tests of linearity of the specification. This is then followed by specifications solely including primary and secondary schools, then results by the individual key stages. I then examine effects in small and large primary schools to try to examine the effect of the direct peer influence, and then examine the effects within key stage within small schools. I follow this up by examining the robustness of the results by comparing them with those from a subset of the sample that only contains schools which appear to have completely random changes in the gender make-up from year to year. I then consider the effects of a measure of poverty for boys and girls. Finally, I repeat the specifications using a value added model to examine the effects of a change in the gender make-up of the peer group on the value added from one key stage to another.

4.6.1 Results in all schools

Table 6 shows regression results for all schools in English, mathematics and science. The initial specification includes all schools and levels, and I estimate equation (7) using analytic weightings, as specified in equation (9):

$$\Delta A_{female,gjc} = \gamma \Delta p_{female,gjc} + \alpha \Delta X_{female,gjc} + \Delta \varepsilon_{female,gjc} \qquad (7)$$

where p is the proportion of pupils in the cohort within the school that is female and X includes year dummies and dummies for the key stage level.

In English, there is a significant negative effect for male pupils of having a more female peer group, whilst for mathematics and science; girls boys experience a significant positive effect of having a more female peer group. Girls appear to be unaffected by having a more female peer group in English. If one considers the effect of the proportion of girls increasing by 10%, then these raw effects would lead to a fall in average English scores for boys by approximately 0.016 standard deviations. For girls, this 10 percentage point increase in the proportion of the peer group that is female would lead to an increase in average mathematics score by 0.007 standard deviations and in science by 0.010 standard deviations. Furthermore, when translated into the effect of a more able peer group, I obtain coefficients that imply that in English, a 1 point increase in the average ability of the peer group is associated with a 0.23 point drop in the individual boys' outcomes, whilst for mathematics, a similar increase in peer ability is associated with a 1.61 point drop in girls average outcomes[27], whilst for science, a 1 point increase in the average ability of the peer group is associated with a 3.4 point decrease in individual girls' outcomes.

In order to check that the model is valid, it is necessary to examine the linearity of the estimates for the coefficient on the proportion of pupils that is female. Figure 9 shows adjusted variable plots for the pooled regressions in English, mathematics and science. In the graphs, the x axis represents the 50 quantiles of

[27] Methodology for calculating this figure, as used by Hoxby (2000) is detailed in the appendix to this chapter.

the proportion of pupils that is female within the cohort in the school and the y axis represents the mean change in average outcomes that can be attributed to just the change in female proportion for each quantile. The fitted line is the fitted regression line. These all appear to follow a fairly linear pattern, other than for females in English, which does not seem to follow any real pattern. However, it is necessary to check this linearity.

Table 7 reports the results of regressions for equation (8) using the pooled specification above, including terms interacting the proportion of pupils that is female with the quartile that it is in. Using an F-test, in all of the subjects for all of the pupils, I do not reject the null that the coefficients on all of these interaction terms are equal, and equal to zero, so I do not reject the null of linearity.

Table 8 shows results of regressions within primary schools, secondary schools and by key stage. Beginning with primary schools, there is a strong significant negative effect on increasing the proportion of pupils within the cohort that is females on male pupils in English. This same effect is seen in both key stage 1 and key stage 2 results, although the magnitude of the coefficient at key stage 1 is much larger than that at key stage 2. For mathematics and science, both male and female pupils see a significant and positive effect of a more female peer group in primary schools. Within secondary schools as a whole and at the finer level of key stages 3 and 4, the only significant effects of a more female peer group are a strongly negative effect for males in English. In terms of the size of effect, at key stage 1, a 10 percent point increase in the proportion of pupil who are female has the following effects; a 0.018 standard deviation fall in boys' average outcomes in English, a 0.011 standard deviation increase in boys' and a 0.007 standard deviation increase in girls' average outcomes in mathematics. For key stage 2, the proportion of girls in the class has no impact on boys' outcomes in English whilst girls' average outcomes are increased by 0.009 standard deviations in English with a 10 percentage point increase in the proportion of the peer group that is female. In mathematics, both boys' and girls' average outcomes are raised by 0.006 and 0.007 standard deviations respectively, whilst for science boys' and girls' outcomes are raised by 0.012 and

0.009 standard deviations respectively. For key stages 3 and 4, a 10 percentage point increase in the proportion of pupils who are female is associated with a 0.033 and 0.028 standard deviation fall in boys' average outcomes in English[28].

Table 9 shows the results translated into the effect of a change in the ability of the peer group based on the change in the gender make-up of the peer group. As with the results presented in Table 6, I find large, negative, effects of an exogenous increase in the peer group ability in mathematics and science at key stages 1 and 2. These effects are of a much larger magnitude than would be credibly expected, and are in the opposite direction to those expected. As with Hoxby (2000), and as suggested by Lavy and Schlosser (2007), the magnitude and direction of these translated effects suggest that something other than merely an ability spillover is occurring. As such, I do not consider these translated effects any further, and merely concentrate on the effect of a 'more female' peer group on outcomes.

4.6.2 Small and large primary schools

Table 10 shows results from the subset of schools that are either defined as small or large schools. The small school definition is a school that does not go over the limit of 30 pupils within the cohort in any year observed in the data, indicating a very high probability that there is only one class within the cohort, and the large school is any school that has more than 30 pupils in all of the observed cohorts, indicating multiple classes. The results for the large primary schools are not significantly different from those for primary schools as a whole. However, within the small primary schools, there is a much larger negative effect on boys in English of a more-female peer group. However, much of the larger magnitude can be explained by the much larger coefficient within key stage 1 scores in small primary schools, which is approximately two and a half times as large as the coefficient for key stage 2. This difference may be explained as results at key stage 1 are generally more noisy than those at other key stages. The only other significant effect within small primary schools is a positive effect

[28] I have not considered here any result that is insignificant at the 10 percent significance level.

for girls in mathematics, which again is being driven by a large effect at key stage 1.

4.6.3 Robustness checks

It is possible that some schools have selection policies based on the gender of pupils, which could affect the results that are gained for the effects of a more female peer group on outcomes. In order to check that the results are not biased by unobserved selection policies, Table 11, Table 12 and Table 13 show comparisons between regressions with all of the schools included, for all of the specifications described above, and a subsample of schools which have apparently random changes in the gender make-up of cohorts. In general, the full sample results do not significantly differ from the random. In Table 11, for English, there are no major differences between the full sample and the apparently random sample. In mathematics, there is a small difference between the results in all levels and schools pooled for males, but this is generally an insignificant difference. Similarly, there is a difference of a reasonably large magnitude between the full sample and credibly random sample for boys in mathematics at key stages 3 and 4, but these are not significant differences. Similarly, there is a large difference between the random and the full sample for key stage 3 for males in science, but again, this is not significant.

Table 13 shows the comparison between the sub-sample of schools that have apparently random changes in gender make-up in the small and large primary schools. There is only one significant difference between the two sets of results, and that is for females in English at key stage 2. However, neither the result in the credibly random sample, nor the result in the full sample is significantly different from 0, so it does not affect my results.

4.6.4 Socioeconomic status

Table 14 shows a breakdown by key stage of results within primary schools using a measure of socioeconomic status of the school, the proportion of male and female pupils who receive free school meals within the cohort. Introducing this measure has no significant effect on the coefficient on the proportion of pupils in the cohort that is female in any subject, in any assessment level. Furthermore, the gender specific socioeconomic status has a significant negative effect on outcomes. For instance, boys in a cohort with a large proportion of males with free school meals do significantly worse in all subjects, and similarly for girls. However, in English at key stage 1, there is a small anomaly. The proportion of male FSM pupils in the cohort in the school actually has a small significant positive impact on females' results. This same effect is seen in Table 15 at key stages 3 and 4 for mathematics, and at key stage 4 for science.

For the gender specific socioeconomic status, the effect seen is constant through primary school, and then increases through secondary school, with the effects for male and female pupils not significantly different.

Since there is no significant change in the coefficient on the proportion of pupils in the school-cohort that is female, I conclude that the socioeconomic status of the school has the same effect on boys and girls and the omission of this variable is not creating any bias in the results.

4.6.5 Value added results

Table 16 shows the results of the estimation of equation (7), with the dependent variable as the average within cohort male (female) value added from one key stage to the next for pupils that stay within the same school, except between key stages 2 and 3, since almost all pupils are registered at a different school between these exams. Beginning by looking at the results for all schools and levels pooled, the only coefficients that are significantly different from zero are for females in mathematics and science. Examining the results at a finer level, it can be seen that this overall result is being driven by a large effect of a more female peer group on value added from key stage 2 to key stage 3, which also drives the large value added observed in the secondary schools for girls in mathematics and science.

In English, a more female peer group has a positive effect on boys and girls at key stage 2. However, comparing the regression results in Table 16 with those from Table 8 it can be seen that this may be due to pupils being disadvantaged at key stage 1 by having a more female peer group, and this disadvantage being reduced over time, with it actually becoming an advantage for girls. However, any advantage gained by girls from having a more female peer group from key stage 2 to key stage 3 seems to be eliminated between key stages 3 and 4, with a large significant negative effect on the proportion of pupils that is female. Finally, examining this value added measure in small primary schools, Table 17 shows the results and there is no observed significant effect of a more female peer group on this value added measure at key stage 2 in small schools

4.6.6 Validity of results.

The identification strategy utilised in this chapter depends on year-on-year changes in the gender make-up of the cohort being random. However, having observed the gender make-up of the cohort, it is possible that a parent may decide to move their child to a different school based on the proportion of pupils who are female, which could result in non-random changes in the school gender mix. In order to check this possibility, I use a pupil level dataset and examine the gender make-up of a child's peers at age 7, and whether this is correlated with the decision to change school between key stage 1 and key stage 2. I consider a peer group at age 7 to be male-dominated if more than 60% of the peer group is male, and similarly a female dominated peer group to be one with 60% or more female. Some schools which enter pupils for both key stage 1 and key stage 2 examinations admit more pupils into year 3 by creating new classes. In order to remove these from the sample, I only consider schools where 80% or more of the pupils at key stage 2 were also in the school at key stage 1. For boys, the correlation with having a male dominated peer group at age 7 and being in the same school at age 11 is -0.0008 (P=0.5381), whilst for girls, the correlation is -0.0042 (P=0.0008). So for girls, there is a significant, but very small correlation between being in a male dominated cohort at age 7 and changing schools for key stage 2. For female dominated peer groups, for boys, the correlation between having as female dominated peer group and being in the same school at age 11 is 0.0064 (P=0.0000), whilst for girls it is 0.0053 (P=0.0000). So, again there is a significant, but very small correlation between being in a female-dominated peer group and staying in the same school. However, whilst there is a statistically significant correlation, the correlation is not quantitatively significant and since they are unlikely to have overly affected the results.

4.7 Conclusions

In this chapter, I have examined the effect of seemingly exogenous changes in the gender make-up of a child's within-school peer group using year to year changes in the proportion of girls within the school as an explanatory variable for the outcomes at key stage 1, key stage 2, key stage 3 and GCSE.

The results show significant negative effects of a more female peer group on male pupils in English, robust across specifications, and a significant positive effect of a more female peer group in mathematics and science for both males and females in primary schools. Hoxby (2000) uses the effect of a change in the proportion of pupils that is female to try and estimate the effect of a credible change in the ability of the peer group, although she does qualify the results with the proviso that her results are of a too high a magnitude to be plausible. However, considering the results I obtain here, due to the considerably higher scores in English achieved by female students, and the slightly higher scores in mathematics and science achieved by male students in primary schools (see Table 4), I find large negative effects of a more able peer group for boys in English at all stages of education except key stage 2, and in mathematics and science for both boys and girls in primary schools, in contradiction with current established literature. For example Lefgren (2004b) finds significant positive effects of a more able peer group, a finding that is backed up by Zimmerman (2003).

This is not the whole story. In mathematics and science, the results of male and female pupils is very closely matched, and so a very large change in the gender make-up of the peer group is required in order to have any noticeable change in the ability of the peer group, meaning that the estimates of the effect of a change in the ability of the peer group would require such a large change in the gender make-up of the cohort that we cannot take the implied result seriously.

Lavy and Schlosser (2007) show that if their results looking at the effect of a more female peer group were driven by changes in the ability of the peer group due to this change in the gender mix, then, in contradiction with the established literature, an increase in the ability of the peer group would lead to a decrease in the pupils' outcomes. They argue that there is some other factor, such as behaviour, that affects the students' outcomes. They demonstrate that an increase in the proportion of girls leads to general increases in academic outcomes, and find that the presence of more female peers lowers classroom violence, whilst improving inter-student, and student-teacher relationships. However, this is not attributed to an individual improvement in behaviour, but rather a compositional effect. This would help to explain my results in mathematics and science, but not for male students in English.

The change in the gender make-up of the peer group could have an influence on the behaviour within the classroom. Younger and Warrington (1996) consider the interactions within the classrooms and the behaviour associated with boys and girls in the classroom. For boys there is an apparent stigma associated with working hard. Furthermore, there is also evidence that boys require more behavioural management than girls. According to the data in PLASC, 70.9 percent of children with statements of special educational needs are boys. This is further shown by the fact that 5 times as many boys are permanently excluded from schools than girls. However, these figures may be slightly misleading, as it has been conjectured that there has been an over-identification of special educational needs (SEN) in boys and a similar under-identification of SEN in girls. In addition, Francis (2000) concludes that boys tend to be louder and more demanding within the classroom, but rather than this directly hindering the boys' own outcomes, it may be having a detrimental effect on all of their classroom peers.

Whilst not being affected directly by their peers, the gender make-up of classrooms may lead to differential teaching methods within the classroom. Whilst teachers may believe that they do not use different methods with girls and boys, Younger et al (1999) find evidence that boys and girls are treated very differently in the classroom. Students claim that boys receive more negative attention than the girls, and there is evidence that teachers have a lower tolerance level to boys' behaviour than to girls, which can "lead to male disillusionment and a negative reaction to learning". (Younger et al (1999), 339) However, they also comment that there is little evidence in observed lessons that boys are given "more support than girls in the teacher-learning process" (Younger et al (1999), 339). Furthermore, Dee (2007) finds that girls taught in a classroom with a female teacher and boys taught with a male teacher tend to perform better than pupils with a teacher of the opposite gender, suggesting that female teachers may direct learning in a way that is more likely to benefit girls rather than boys. This, when combined with Macleod (2005) who comments that only 15.7% of all primary school teachers in England are male and half of 5 to 11 year olds have no contact with male teachers implies that girls are likely to benefit more in education due to the gender of teachers.

Considering the difference between the single classroom cohorts in primary schools with the full sample, there is a much larger magnitude negative effect within the single classroom case for boys in English, which tends to lead us towards the conclusion of more behavioural issues with boys, or possibly the impact of a more female orientated teaching method, leading to disadvantages for boys. Further, it appears that girls benefit from an environment more suitable for learning in mathematics if there are more girls in the classroom, whether through better behaviour or more directed teaching. This model has less noise in it than the larger schools, as I can observe directly the within classroom peer group. In the large schools[29], the negative effects for boys in English disappear, but apparent positive effects are seen in mathematics and science which are not seen in the small schools' case.

---

[29] Schools with more than 30 pupils in the cohort.

Overall, the results imply that in primary schools at least, boys would benefit greatly from being taught English in single sex classes, which would have little effect on girls' outcomes, whilst in mathematics and science, different policies would benefit boys and girls: boys would be better off in a more female classroom, whilst girls would be better off in an all female classroom. This last conclusion, however, is in contradiction to Smithers and Robinson (2006), who claim there is little evidence to suggest that there is no advantage of teaching girls in a single sex environment, contradicting the long-held view that in schools girls are distracted by boys in the classroom, and other arguments that girls and boys brains develop differently and thus require different emphases in teaching. Their research examines data from across the world, and finds little impact of consistent superior performance in single sex schools, and whilst single schools may appear to perform better, it is not a function of the fact that they are single sex.

Smithers and Robinsons' (2006) research is almost in direct contradiction to Younger et al (2005) who when examining whole school approaches to raising boys' achievement consider the effects of single-sex classes. They find evidence that "girls and boys feel more at ease in such classes, feel more able to interact with learning and to show real interest without inhibition and often achieve more highly as a result"[30] Thus, whilst my results back up Younger et al (2005) for English and for girls in mathematics and science, single sex classes in mathematics and science for boys would have a detrimental effect. Furthermore, Jackson (2002) finds that single sex classes are likely to have positive effects for girls, but male only classes may exacerbate the macho male cultures inherent in schools.

Since the methodology used here takes advantage of variation in the proportion of the cohort that is female, it is not possible to make definitive conclusions of the effect of educating pupils in single sex classrooms. However, the results obtained suggest that boys would benefit at all ages from being taught English in English schools with as small a proportion of girls as possible. In mathematics

---

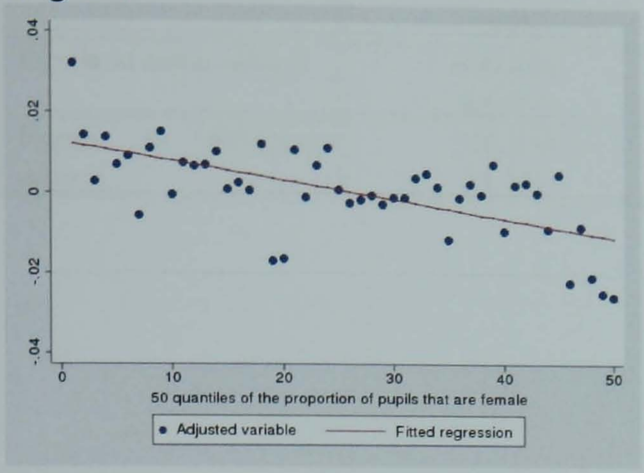
[30] Younger et al (2005) page 12

and science, the results shown here tend to imply that both boys and girls benefit from having more girls in the classroom. However, it is not possible to increase the proportion of girls for both boys and girls, implying that a mix of the genders is optimal in both mathematics and science.

Figure 9 **Adjusted variable plots of the pooled regressions.**

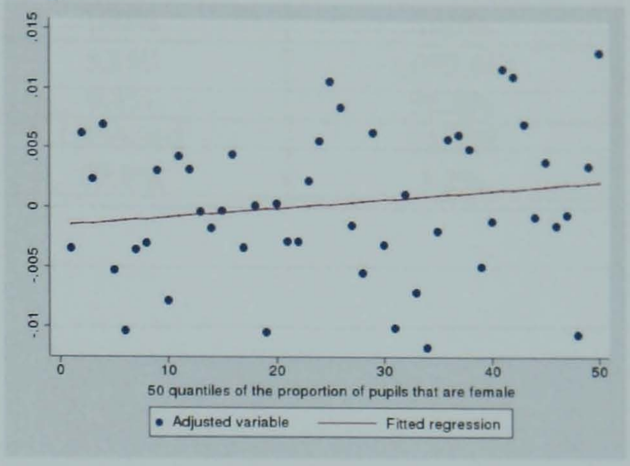
Male

English

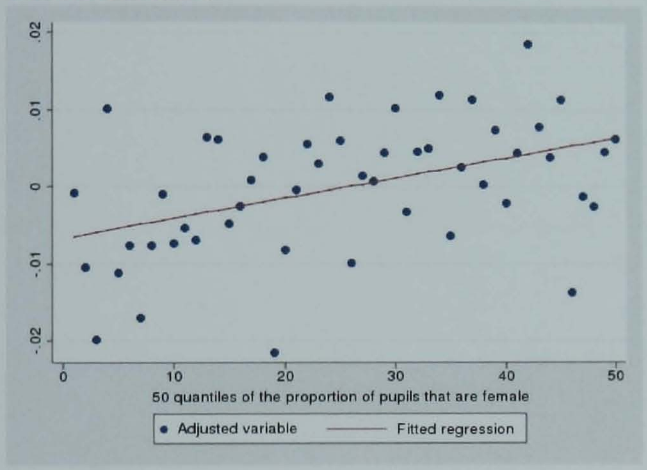


Female

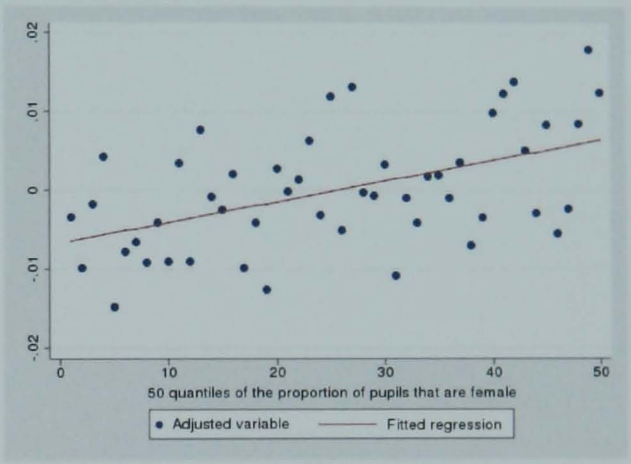
English



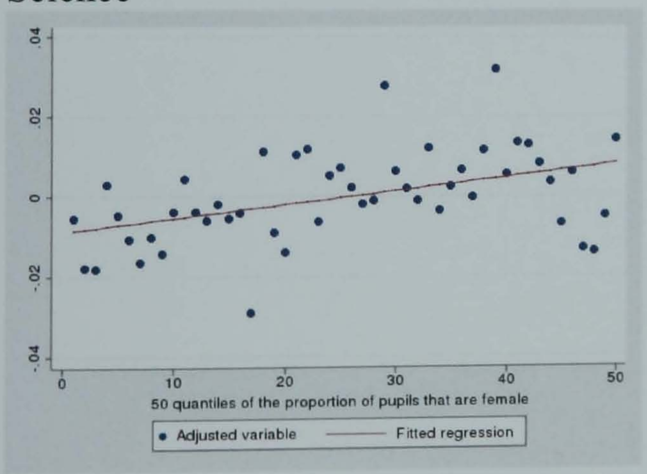
Mathematics



Mathematics



Science



Science

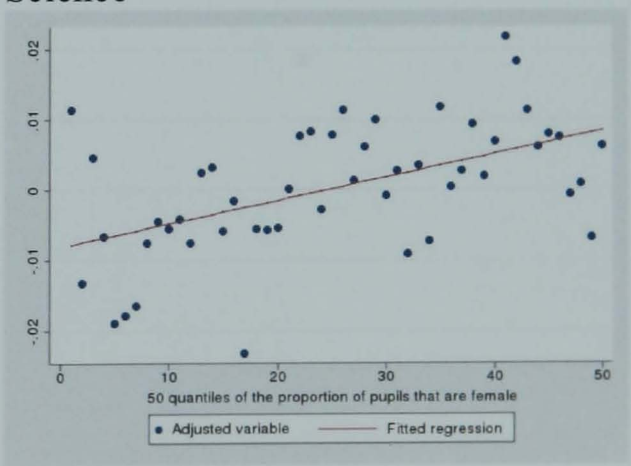


Table 3 Proportion of pupils that stay at the same school between key stages

	Key Stage 1 – 2	Key Stage 2 – 3	Key Stage 3 – 4
Total	911,470 100%	1,359,182 100%	1,066,189 100%
Pupils at same school	600,814 65.9%	5,380 0.4%	1,032,461 96.8%
Pupils at different school	310,656 34.1%	1,353,802 99.6%	33,728 3.2%

Table 4 Summary statistics

	Mean score for males in English	Mean scores for females in English	Mean scores for males in mathematics	Mean scores for females in mathematics	Mean scores for Males in science	Mean scores for females in Science	Proportion of the cohort that is female	Size of cohort within school
Pooled Specification								
Mean	-0.174	0.182	0.006	-0.006	0.002	-0.002	0.488	112.566
Standard Deviation	(0.419)	(0.400)	(0.408)	(0.404)	(0.431)	(0.439)	(0.072)	(84.753)
Observations	160604	160604	160604	160604	89344	89344	160604	160604
Key Stage 1								
Mean	-0.156	0.164	0.005	-0.004			0.489	39.350
Standard Deviation	(0.432)	(0.401)	(0.434)	(0.408)			(0.090)	(19.030)
Observations	71260	71260	71260	71260			71260	71260
Key Stage 2								
Mean	-0.150	0.154	0.032	-0.033	0.007	-0.007	0.492	41.892
Standard Deviation	(0.444)	(0.417)	(0.430)	(0.426)	(0.491)	(0.495)	(0.088)	(21.502)
Observations	72248	72248	72248	72248	72248	72248	72248	72248
Key Stage 3								
Mean	-0.192	0.205	0.008	-0.008	0.017	-0.018	0.483	186.759
Standard Deviation	(0.398)	(0.399)	(0.372)	(0.373)	(0.404)	(0.404)	(0.050)	(58.550)
Observations	10415	10415	10415	10415	10415	10415	10415	10415
Key Stage 4								
Mean	-0.191	0.198	-0.014	0.015	-0.014	0.015	0.488	182.263
Standard Deviation	(0.403)	(0.385)	(0.399)	(0.407)	(0.406)	(0.421)	(0.050)	(61.344)
Observations	6681	6681	6681	6681	6681	6681	6681	6681

Notes: Unit of comparison is the within school , within key stage cohort. The summary statistics for the mean scores at the key stage are generate using weighted values as described in the methodology, whilst those for the proportion of the cohort that is female and the size of cohort within the school are weighted using the inverse of the number of cohorts within schools observed at each key stage. Key stage 1 is not formally assessed for science.

Table 5 **Summary statistics – small primary schools broken down by key stage.**

	Mean score for males in English	Mean scores for females in English	Mean scores for males in mathematics	Mean scores for females in mathematics	Mean scores for males in science	Mean scores for females in Science	Proportion of the cohort that is female	Size of cohort within school
Key Stage 1 in small primary schools								
Mean	-0.129	0.198	0.035	0.023			0.490	22.085
Standard Deviation	(0.476)	(0.443)	(0.481)	(0.453)			(0.111)	(5.545)
Observations	16681	16681	16681	16681			16681	16681
Key Stage 2 in small primary schools								
Mean	-0.097	0.220	0.083	0.032	0.039	0.034	0.494	22.070
Standard Deviation	(0.503)	(0.473)	(0.490)	(0.481)	(0.543)	(0.548)	(0.111)	(5.334)
Observations	13408	13408	13408	13408	13408	13408	13408	13408

Notes: Unit of comparison is the within school , within key stage cohort. The summary statistics for the mean scores at the key stage are generate using weighted values as described in the methodology, whilst those for the proportion of the cohort that is female and the size of cohort within the school are weighted using the inverse of the number of cohorts within schools observed at each key stage. Key stage 1 is not formally assessed for science. A small primary school is defined as one that has 30 or fewer pupils in all observed years.

Table 6 Results for all schools and levels pooled

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
All levels and schools pooled						
Proportion of the within-school cohort that is female	-0.066***	0.011	0.020	0.030**	0.025	0.043**
	(0.013)	(0.013)	(0.012)	(0.012)	(0.018)	(0.018)
Observations	137083	137083	137083	137083	76003	76003
Adjusted R-squared	0.02	0.02	0.04	0.05	0.08	0.11
Estimate of the effect of a 1 point increase in the peer ability due to the change in the gender make-up.	-0.227	0.037	-1.077	-1.611	-1.992	-3.385

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Year and exam dummies are also included. Standard errors are clustered at school level.

Table 7 Testing the linearity of the pooled regressions

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Proportion of pupils that is female (1)	-0.061***	0.018	0.037*	0.037*	0.071**	0.080***
	(0.022)	(0.022)	(0.021)	(0.021)	(0.031)	(0.031)
(1) interacted with 2nd quartile	0.054	0.052	-0.086	-0.088	0.080	-0.034
dummy (2)	(0.080)	(0.078)	(0.066)	(0.062)	(0.086)	(0.084)
(1) interacted with 3rd quartile	0.004	-0.055	-0.022	-0.051	-0.196*	-0.134
dummy (3)	(0.096)	(0.095)	(0.081)	(0.081)	(0.108)	(0.108)
(1) interacted with 4th quartile	-0.020	-0.015	-0.022	0.004	-0.085*	-0.054
dummy (4)	(0.037)	(0.036)	(0.033)	(0.033)	(0.050)	(0.049)
Observations	137083	137083	137083	137083	76003	76003
R-squared	0.02	0.01	0.03	0.03	0.07	0.09
P> F test statistic (2)=(3)=(4)=0	0.8220	0.8999	0.3610	0.3240	0.2831	0.4906

Notes: Robust standard errors in parentheses *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Year and exam dummies are also included. Standard errors are clustered at school level. Quartile dummies are based on 4 quantiles based on the proportion of pupils who are female.

Table 8 The effect of a more female peer group, broken down by primary and secondary schools and key stages

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Pooled primary schools						
Proportion of the within-school	-0.050***	0.008	0.040***	0.030***	0.061***	0.048***
cohort that is female	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.016)
Observations	124297	124297	124297	124297	63217	63217
Adjusted R-squared	0.03	0.02	0.04	0.05	0.10	0.13
Pooled secondary schools						
Proportion of the within-school	-0.131***	0.018	-0.056	0.034	-0.049	0.028
cohort that is female	(0.050)	(0.049)	(0.042)	(0.043)	(0.044)	(0.043)
Observations	12786	12786	12786	12786	12786	12786
Adjusted R-squared	0.02	0.01	0.05	0.05	0.07	0.09
Key Stage 1						
Proportion of the within-school	-0.077***	-0.022	0.049***	0.028*	**N/A**	**N/A**
cohort that is female	(0.015)	(0.014)	(0.016)	(0.016)		
Observations	61080	61080	61080	61080		
Adjusted R-squared	0.01	0.02	0.05	0.05		
Key Stage 2						
Proportion of the within-school	-0.024	0.041***	0.025*	0.029**	0.061***	0.048***
cohort that is female	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Observations	63217	63217	63217	63217	63217	63217
Adjusted R-squared	0.05	0.04	0.06	0.08	0.10	0.13
Key Stage 3						
Proportion of the within-school	-0.137**	0.059	-0.056	0.033	-0.038	0.028
cohort that is female	(0.068)	(0.072)	(0.037)	(0.037)	(0.043)	(0.042)
Observations	8332	8332	8332	8332	8332	8332
Adjusted R-squared	0.03	0.01	0.10	0.13	0.20	0.22
Key Stage 4						
Proportion of the within-school	-0.129**	-0.017	-0.070	0.019	-0.079	0.012
cohort that is female	(0.063)	(0.059)	(0.060)	(0.060)	(0.063)	(0.061)
Observations	4454	4454	4454	4454	4454	4454
Adjusted R-squared	0.01	0.02	0.06	0.04	0.01	0.01

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Year and exam dummies are also included. Standard errors are clustered at school level.

Table 9 **Estimated effect of a 1 point average increase in the ability of the peer group from the change in the gender make-up**

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
Pooled primary schools	**-0.177**	0.027	**-1.120**	**-0.859**	**-5.912**	**-4.663**
	124297	124297	124297	124297	63217	63217
Pooled Secondary Schools	**-0.420**	0.059	1.669	-1.031	1.346	-0.765
	12786	12786	12786	12786	12786	12786
Key Stage 1	**-0.261**	-0.076	**-18.030**	**-10.201**		
	61080	61080	61080	61080		
Key Stage 2	-0.087	**0.149**	**-0.352**	**-0.416**	**-5.912**	**-4.663**
	63217	63217	63217	63217	63217	63217
Key Stage 3	**-0.459**	0.199	0.960	-0.567	0.658	-0.480
	8332	8332	8332	8332	8332	8332
Key Stage 4	**-0.381**	-0.051	6.983	-1.926	4.160	-0.650
	4454	4454	4454	4454	4454	4454

Notes: The coefficients estimated here are the estimated effect of a 1 point change in the ability of the peer group related to a change in the gender make-up of the peer effect. Coefficient in bold indicates that the corresponding estimate of a more female peer group is significantly different from 0 at the 10% significance level. Method of estimation is as per Hoxby (2000). See appendix of chapter 4 for details.

Table 10 Results in the subset of small and large primary schools

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
Large Primary Schools						
Proportion of the within-school	-0.032*	0.006	0.054***	0.021	0.079***	0.065***
cohort that is female	(0.017)	(0.016)	(0.017)	(0.017)	(0.024)	(0.025)
Observations	55211	55211	55211	55211	29099	29099
Adjusted R-squared	0.03	0.03	0.06	0.07	0.13	0.16
Small Primary Schools						
Proportion of the within-school	-0.105***	0.015	0.014	0.056**	-0.020	0.007
cohort that is female	(0.023)	(0.022)	(0.024)	(0.023)	(0.036)	(0.034)
Observations	26030	26030	26030	26030	11732	11732
Adjusted R-squared	0.01	0.01	0.02	0.03	0.05	0.07
Key Stage 1 in small primary schools						
Proportion of the within-school	-0.145***	-0.005	0.022	0.067**	**N/A**	**N/A**
cohort that is female	(0.029)	(0.027)	(0.032)	(0.030)		
Observations	14298	14298	14298	14298		
Adjusted R-squared	0.01	0.01	0.03	0.03		
Key Stage 2 in small primary schools						
Proportion of the within-school	-0.055	0.040	0.001	0.041	-0.020	0.007
cohort that is female	(0.036)	(0.035)	(0.035)	(0.033)	(0.036)	(0.034)
Observations	11732	11732	11732	11732	11732	11732
Adjusted R-squared	0.03	0.03	0.03	0.04	0.05	0.07

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. In square brackets are the translated effects of the coefficients of the exogenous change in peer tests scores that occurs from a change in the gender make-up of the peer group. Method is weighted least squares. A small primary school is defined as one that is observed to have cohort sizes smaller, or equal, than 30 for every cohort observed in the data. A large primary school is defined as one that is observed to have cohort sizes larger than 30 for all of the cohorts observed in the data. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 11 Comparisons of full sample and credibly random sample

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
All levels and schools pooled						
Schools that have apparent random changes in gender make-up	-0.068***	0.014	-0.002	0.035**	0.027	0.063***
	(0.018)	(0.017)	(0.017)	(0.016)	(0.025)	(0.024)
Observations	72684	72684	72684	72684	40176	40176
All Schools	-0.066***	0.011	0.020	0.030**	0.025	0.043**
	(0.013)	(0.013)	(0.012)	(0.012)	(0.018)	(0.018)
Observations	137083	137083	137083	137083	76003	76003
Primary Schools						
Schools that have apparent random changes in gender make-up	-0.053***	0.005	0.029*	0.037**	0.070***	0.066***
	(0.015)	(0.014)	(0.015)	(0.014)	(0.021)	(0.021)
Observations	65870	65870	65870	65870	33362	33362
All Schools	-0.050***	0.008	0.040***	0.030***	0.061***	0.048***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.016)	(0.016)
Observations	124297	124297	124297	124297	63217	63217
Secondary Schools						
Schools that have apparent random changes in gender make-up	-0.124*	0.041	-0.111**	0.034	-0.055	0.051
	(0.065)	(0.063)	(0.055)	(0.055)	(0.058)	(0.055)
Observations	6814	6814	6814	6814	6814	6814
All Schools	-0.131***	0.018	-0.056	0.034	-0.049	0.028
	(0.050)	(0.049)	(0.042)	(0.043)	(0.044)	(0.043)
Observations	12786	12786	12786	12786	12786	12786

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 12 **Comparisons of full sample and credibly random sample, broken down by key stage**

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Key Stage 1						
Schools that have apparent random	-0.081***	-0.019	0.032	0.043**	**N/A**	**N/A**
changes in gender make-up	(0.020)	(0.018)	(0.021)	(0.020)		
Observations	32508	32508	32508	32508		
All Schools	-0.077***	-0.022	0.049***	0.028*	**N/A**	**N/A**
	(0.015)	(0.014)	(0.016)	(0.016)		
Observations	61080	61080	61080	61080		
Key Stage 2						
Schools that have apparent random	-0.023	0.031	0.024	0.030	0.070***	0.066***
changes in gender make-up	(0.021)	(0.021)	(0.020)	(0.020)	(0.021)	(0.021)
Observations	33362	33362	33362	33362	33362	33362
All Schools	-0.024	0.041***	0.025*	0.029**	0.061***	0.048***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Observations	63217	63217	63217	63217	63217	63217
Key Stage 3						
Schools that have apparent random	-0.147*	0.091	-0.105**	0.008	-0.084	0.015
changes in gender make-up	(0.089)	(0.093)	(0.049)	(0.048)	(0.057)	(0.056)
Observations	4456	4456	4456	4456	4456	4456
All Schools	-0.137**	0.059	-0.056	0.033	-0.038	0.028
	(0.068)	(0.072)	(0.037)	(0.037)	(0.043)	(0.042)
Observations	8332	8332	8332	8332	8332	8332
Key Stage 4						
Schools that have apparent random	-0.111	0.002	-0.139*	0.045	-0.057	0.070
changes in gender make-up	(0.083)	(0.078)	(0.081)	(0.080)	(0.082)	(0.080)
Observations	2358	2358	2358	2358	2358	2358
All Schools	-0.129**	-0.017	-0.070	0.019	-0.079	0.012
	(0.063)	(0.059)	(0.060)	(0.060)	(0.063)	(0.061)
Observations	4454	4454	4454	4454	4454	4454

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Key stage 1 is not formally assessed for science Standard errors are clustered at school level

Table 13 Comparisons of full sample and credibly random sample in small and large primary schools

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
Large Primary Schools						
Schools that have apparent random changes in gender make-up	-0.020	0.034	0.053**	0.045**	0.092***	0.098***
	(0.022)	(0.021)	(0.023)	(0.022)	(0.032)	(0.033)
	29627	29627	29627	29627	15491	15491
All Schools	-0.032*	0.006	0.054***	0.021	0.079***	0.065***
	(0.017)	(0.016)	(0.017)	(0.017)	(0.024)	(0.025)
	55211	55211	55211	55211	29099	29099
Small Primary Schools						
Schools that have apparent random changes in gender make-up	-0.117***	-0.024	-0.005	0.033	-0.036	-0.010
	(0.031)	(0.030)	(0.032)	(0.029)	(0.047)	(0.045)
	13813	13813	13813	13813	6181	6181
All Schools	-0.105***	0.015	0.014	0.056**	-0.020	0.007
	(0.023)	(0.022)	(0.024)	(0.023)	(0.036)	(0.034)
	26030	26030	26030	26030	11732	11732
Key Stage 1 in small primary schools						
Schools that have apparent random changes in gender make-up	-0.153***	-0.018	-0.007	0.061	N/A	N/A
	(0.039)	(0.036)	(0.042)	(0.040)		
	7632	7632	7632	7632		
All Schools	-0.145***	-0.005	0.022	0.067**	N/A	N/A
	(0.029)	(0.027)	(0.032)	(0.030)		
	14298	14298	14298	14298		
Key Stage 2 in small primary schools						
Schools that have apparent random changes in gender make-up	-0.070	-0.034	-0.002	-0.006	-0.036	-0.010
	(0.050)	(0.049)	(0.047)	(0.043)	(0.047)	(0.045)
	6181	6181	6181	6181	6181	6181
All Schools	-0.055	0.040	0.001	0.041	-0.020	0.007
	(0.036)	(0.035)	(0.035)	(0.033)	(0.036)	(0.034)
	11732	11732	11732	11732	11732	11732

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. A small primary school is defined as one that is observed to have cohort sizes smaller, or equal, than 30 for every cohort observed in the data. A large primary school is defined as one that is observed to have cohort sizes larger than 30 for all of the cohorts observed in the data. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 14 OLS estimation including socioeconomic factors within primary schools

	English		Mathematics		Science	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Key Stage 1						
Proportion of the within-school cohort that is female	-0.073***	-0.025*	0.052***	0.026*	N/A	N/A
	(0.014)	(0.013)	(0.016)	(0.015)		
Proportion of males that receive FSM within cohort	-0.465***	0.023*	-0.389***	0.012		
	(0.015)	(0.014)	(0.017)	(0.016)		
Proportion of females that receive FSM within cohort	0.021	-0.415***	0.013	-0.357***		
	(0.014)	(0.013)	(0.016)	(0.016)		
Observations	61080	61080	61080	61080		
Adjusted R-squared	0.03	0.04	0.06	0.06		
Key Stage 2						
Proportion of the within-school cohort that is female	-0.019	0.035**	0.029*	0.024	0.065***	0.042***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Proportion of males that receive FSM within cohort	-0.434***	-0.015	-0.374***	0.009	-0.370***	0.009
	(0.016)	(0.015)	(0.016)	(0.015)	(0.017)	(0.016)
Proportion of females that receive FSM within cohort	0.026*	-0.399***	0.007	-0.376***	0.014	-0.386***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Observations	63217	63217	63217	63217	63217	63217
Adjusted R-squared	0.07	0.05	0.07	0.09	0.11	0.14

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. In square brackets are the translated effects of the coefficients of the exogenous change in peer tests scores that occurs from a change in the gender make-up of the peer group. Method is weighted least squares. FSM is free school meals. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 15 OLS estimation including socioeconomic factors within secondary schools

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
Key Stage 3						
Proportion of the within-school cohort that is female	-0.124*	0.062	-0.044	0.037	-0.026	0.030
	(0.068)	(0.072)	(0.037)	(0.037)	(0.042)	(0.041)
Proportion of males that receive FSM within cohort	-0.524***	-0.044	-0.488***	-0.080**	-0.490***	-0.028
	(0.067)	(0.072)	(0.039)	(0.038)	(0.043)	(0.043)
Proportion of females that receive FSM within cohort	0.072	-0.433***	0.018	-0.495***	0.001	-0.572***
	(0.065)	(0.073)	(0.036)	(0.036)	(0.042)	(0.042)
Observations	8332	8332	8332	8332	8332	8332
Adjusted R-squared	0.04	0.02	0.12	0.15	0.21	0.25
Key Stage 4						
Proportion of the within-school cohort that is female	-0.130**	-0.020	-0.072	0.017	-0.079	0.010
	(0.062)	(0.058)	(0.059)	(0.059)	(0.061)	(0.061)
Proportion of males that receive FSM within cohort	-0.749***	-0.051	-0.658***	-0.137**	-0.673***	-0.135**
	(0.070)	(0.064)	(0.065)	(0.062)	(0.069)	(0.068)
Proportion of females that receive FSM within cohort	0.033	-0.513***	-0.046	-0.590***	-0.030	-0.596***
	(0.064)	(0.060)	(0.062)	(0.059)	(0.066)	(0.066)
Observations	4454	4454	4454	4454	4454	4454
Adjusted R-squared	0.04	0.04	0.09	0.07	0.03	0.03

Notes: Dependent variable is the change in mean key stage score within school cohort for male (female) pupils. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. In square brackets are the translated effects of the coefficients of the exogenous change in peer tests scores that occurs from a change in the gender make-up of the peer group. Method is weighted least squares. FSM is free school meals. Each cell represents a separate regression. Key stage 1 is not formally assessed for science. Standard errors are clustered at school level.

Table 16 Value added

	English		Mathematics		Science	
	Male	Female	Male	Female	Male	Female
All levels and schools pooled						
Proportion of the within-school cohort	-0.016	0.014	0.016	0.067***	-0.042	0.066*
that is female	(0.028)	(0.028)	(0.022)	(0.023)	(0.038)	(0.037)
Observations	25908	25908	25908	25908	12306	12306
Adjusted R-squared	0.06	0.03	0.13	0.20	0.33	0.43
Pooled Secondary Schools						
Proportion of the within-school cohort	-0.101**	-0.033	0.005	0.086***	-0.042	0.066*
that is female	(0.043)	(0.044)	(0.027)	(0.031)	(0.038)	(0.037)
Observations	12306	12306	12306	12306	12306	12306
Adjusted R-squared	0.06	0.03	0.22	0.33	0.33	0.43
Key Stage 2						
Proportion of the within-school cohort	0.089***	0.071**	0.031	0.044		
that is female	(0.031)	(0.030)	(0.033)	(0.033)		
Observations	13602	13602	13602	13602		
Adjusted R-squared	0.03	0.01	0.06	0.06		
Key Stage 3						
Proportion of the within-school cohort	-0.084	0.054	-0.011	0.108***	-0.075	0.086*
that is female	(0.070)	(0.073)	(0.036)	(0.037)	(0.051)	(0.049)
Observations	8284	8284	8284	8284	8284	8284
Adjusted R-squared	0.08	0.04	0.27	0.41	0.38	0.46
Key Stage 4						
Proportion of the within-school cohort	-0.119	-0.194*	0.033	0.016	0.002	0.010
that is female	(0.098)	(0.101)	(0.048)	(0.045)	(0.058)	(0.060)
Observations	4022	4022	4022	4022	4022	4022
Adjusted R-squared	0.02	0.03	0.08	0.11	0.22	0.34

Notes: Dependent variable is the change in mean value added score from one key stage to the next within school cohort for male (female) pupils. Only pupils who remain in the same school from key stage 1 to key stage 2 and key stage 3 to key stage 4 are included, whilst pupils who change schools between key stage 3 and 4 are included. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. Each cell represents a separate regression. Science has no regressions at key stage 2 as the pupils are not formally assessed at key stage 1 for science. Standard errors are clustered at school level

Table 17 Value added in small primary schools

	English		Mathematics	
	Male	**Female**	**Male**	**Female**
Small school Key Stage 2				
Proportion of the within-school cohort that is female	0.072	0.017	-0.055	0.011
	(0.064)	(0.059)	(0.067)	(0.058)
Observations	2848	2848	2848	2848
Adjusted R-squared	0.03	0.01	0.05	0.04
Large school Key Stage 2				
Proportion of the within-school cohort that is female	0.103	0.085	0.103	0.085
	(0.064)	(0.054)	(0.064)	(0.054)
Observations	4938	4938	4938	4938
Adjusted R-squared	0.03	0.01	0.03	0.01

Notes: Dependent variable is the change in mean value added score from one key stage to the next within school cohort for male (female) pupils. Only pupils who remain in the same school from key stage 1 to key stage 2 and key stage 3 to key stage 4 are included, whilst pupils who change schools between key stage 3 and 4 are included. Robust standard errors in parentheses. *** denotes significance at the 1% level, ** denotes significance at the 5% level, * denotes significance at the 1% level. Method is weighted least squares. Each cell represents a separate regression. A small primary school is defined as one that is observed to have cohort sizes smaller, or equal, than 30 for every cohort observed in the data. A large primary school is defined as one that is observed to have cohort sizes larger than 30 for all of the cohorts observed in the data. Each cell represents a separate regression. Science has no regressions at key stage 2 as the pupils are not formally assessed at key stage 1 for science. Standard errors are clustered at school level

A4.8 Estimating the effect of a change in the ability of the peer group.

Measuring the effect of peer ability.

The three mechanisms for peer effects are discussed above. We wish to examine the apparent effect of the peer group ability on outcomes. In order to do this, we need to begin with the assumption that all of the gender peer effects are due to the change in ability of the peer group.

In order to examine this, we first need to establish the average overall outcome which is not affected by our peer ability measure, which we can calculate by using the estimated coefficient from the proportion of pupils that is female.

$$\hat{\mu}_{female,g} = weightedmean(A_{female,gjc} - \hat{\gamma}_{female,g} p_{female,gjc})$$

Now, if females on average score 1 point higher than males, then an increase in the proportion of pupils that is female of 10% would lead to an increase in peer ability of 0.1 points.

Thus, we can calculate the percentage change in the proportion of pupils that is female required to produce a 1 point increase in the peer group ability.

$$change = \frac{1}{\hat{\mu}_{female,g} - \hat{\mu}_{male,g}}$$

Thus, in order to calculate the effect of a change in peer group ability of 1 point, we multiply the change by the coefficient on the proportion of pupils that is female.[31]

[31] This methodology is the same as that employed in Hoxby (2000)

Chapter 5 Peer Effects In English Secondary Schools.[32]

A study of the effect of a more able peer group on outcomes at age 16 using credibly random allocation to classrooms.

5.1 Introduction

In chapter 4, I have shown that, as suggested by the Coleman Report (1966), a child's peers have a significant effect on their outcomes, but I have so far only isolated the effect of a change in the gender-mix of the peer group. In this chapter, I move on to examine the effect of a more able peer group on pupils' outcomes at age 16.

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. 2007). Whether the sorting of children in these ways has 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 who 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 a more able peer group from unobserved heterogeneity is not straightforward, as discussed in chapter 3, but there has been a rapid

[32] This chapter is co-authored with Adele Atkinson and Professors Simon Burgess Paul Gregg and Carol Propper

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 (2008)

In this chapter I use a unique dataset for England,[33] containing data on the classes in which pupils are taught at ages 15 and 16, giving us data on the peer group that the students directly experience. This dataset contains their class group for lessons up to GCSE, and also their results at key stage 3 and GCSE.[34] From the key stage 3 results, we can construct the mean prior attainment of the peer group. 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 ordering to apparent random allocation. We estimate the effect of a more able peer group using a sample of schools with a credibly random distribution of pupils, and for validation, we utilise a two stage least squares technique developed by Lefgren (2004b). 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 discusses identification issues. Section 3 examines the data. Section 4 discusses the results. Section 5 offers some concluding discussion.

5.2 Methodology

Researchers face substantial problems[35] 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

[33] Data collected to examine the effects of the introduction of performance related pay. See Atkinson et al (2009)
[34] The structure of the English school system is discussed in more detail in chapter 2.
[35] These problems, and possible solutions to the problems are detailed in chapter 3.

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. Also used may be the ability and potential of the child, which may be based on a teacher's assessment of a child rather than merely their performance in examinations.

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), as discussed in chapter 3:

$$GCSE_{ij} = A_t[F_{ij}, S_{ij}, \mu_{ij}, \varepsilon_{ij}] \quad (10)$$

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 be cumulated into a lagged attainment measure this can be re-written as:

$$GCSE_{ijt} = \beta_1 F_{ijt} + \beta_2 S_{ijt} + \lambda A_{ij,t-2} + v_{ij} + \eta_{jit} \quad (11)$$

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

$$S_{ijt} = C_{ijt} + T_{jt} + S_t \qquad (12)$$

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 score of the set, not including the subject child, where key stage 3 is taken two years prior to their GCSE examinations. 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 is 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 follow a similar strategy to Lefgren (2004b); 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 tiers 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 for which the pupils are entered. Due to this tiering policy, it is unlikely that many, if any, schools have a random setting policy for the whole school, but our identification strategy should allow us to use schools with apparently 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

$$GCSE_{ij_t} = \alpha + X_{ij}\delta_j + classave_{ij_{t-2}}\gamma_j + S_t + T_t + \varepsilon_{ij} \tag{13}$$

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 (*S*). These will remove 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 *(T)*.

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 and gender and a ward-level measure of local income 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[36] 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 mathematics. For example, in higher tier mathematics, choosing a cut-off of 0.35 leaves 7 schools in our sample, but if the cut-off were to be reduced to 0.30, then only 3 schools would be included. We can check for any residual bias by comparing the results those for the two stage least squares 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 use 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 mathematics:

$$classave_{ij_{t-2}} = a + X_{ij}\beta_j + R_{sj}^2 KS3_{ij_{t-2}}\psi_j + S_t + T_t + u_{ij} \quad (14)$$

$$GCSE_{ij_t} = \alpha + X_{ij}\delta_j + classave_{ij_{t-2}}\gamma_j + S_t + T_t + \varepsilon_{ij} \quad (15)$$

[36] 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.[37]

This two stage least squares 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.

5.3 Data

We use a unique dataset 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. (2009)). 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.

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

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 mathematics. 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 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 the pupils' home postcode allows us to map the ward level index of deprivation. The index we use is the index of income deprivation from the 2000 Indices of Deprivation from the Department of the Environment, Transport and Regions[38] of local area deprivation. This measure of income deprivation gives a measure of the level to which people within the ward have low incomes. Hence, we have some idea of the demographics of the area in which the pupils live, and thus also some of the characteristics of the pupils and their neighbourhood.

This dataset has data on pupils results at ages 14 and 16 (key stage 3 and GCSE respectively) in English mathematics and science.[39] 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*. For the sake of the analysis here, 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.[40] 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.

---

[38] This index of local deprivation is available from http://www.communities.gov.uk/archived/publications/citiesandregions/indicesdeprivation , accessed 24/09/2009

[39] The structure of examinations in England is discussed in more detail in chapter 3.

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

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 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, in some classrooms, the teaching may not be focussed on one single tier, but instead focussing teaching for pupils of mixed ability. These mixed ability groups will have 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.

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. However, these mixed tier entry classrooms could have different characteristics, ranging from a generally very good classroom with a few low achievers to a generally very bad classroom with a few high achievers. Necessarily, these types of classes are likely to have different focus on the syllabus content, so identifying the effects of a more able peer group would be a complex task. Due to the complexity of these mixed tier entry classrooms, we do not consider mixed tier entry classes in this analysis.

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 18 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 19 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 mathematics. 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" (Atkinson et al (2004, 21)).

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, GCSEs are examined in tiers, so 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 19. 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 10 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 11 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. We 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.[41] Furthermore, we include all of the subject specific key stage 3 scores as individual dummies.[42] 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.

5.4 Results

5.4.1 OLS estimates

Table 21 contains OLS estimates of the classroom level peer effects present for English and mathematics. 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 2) and in column 3 prior attainment in the other key stage 3 subject is also included. Column 4 introduces school fixed effects so that we are estimating within schools and finally Column 5 introduces teacher fixed effects. In Column 5 those teacher or teacher combinations that appear only once and retained in the sample and in Columns 6 and 7 we repeat columns 4 and 5 but only include observations where the teacher is observed teaching at least two classes.

5.4.2 Pooled estimation

Starting with the results for English and in the upper panel of Table 21, using specification 1, 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

[41] 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.

[42] 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

average lagged attainment changes by one grade,[43] a result is seen equivalent to raising a pupil's attainment by between 1.17 GCSE grades. Alternatively, a 1 standard deviation increase in the peer ability measure leads to a 0.69 standard deviation increase in GCSE score. The rows reflect the impact of moving to within tier estimation for English in the upper panel and mathematics 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 for 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 key stage 3 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 4 and 5 suggest that an increase in average peer ability by 1 standard deviation at key stage 3 in English raises pupil attainment by 0.4 GCSE grades or approximately 0.26 standard deviations. Thus these estimated effects of peer effects within the classroom are moderately large. The picture for mathematics is broadly the same except that the estimated coefficients are somewhat higher with conditional estimates of around 0.6, or alternatively a 1 standard deviation increase in the peer ability measure leading to 0.67 grades at GCSE, or a 0.37 standard deviation increase in individuals' outcomes.

[43] This change is roughly a change of 1.09 standard deviations of the class average score in English, and 0.87 standard deviations for mathematics. Standard deviation of class average in English is 0.917, whilst for mathematics it is 1.149

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.

5.4.3 Within tier estimation

As noted earlier, setting is very common at school level especially in mathematics. This is, in part, to 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 key stage 3 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 within the higher tier, with an increase of the peer ability by 1 standard deviation[44], within the tier, raises the outcome at GCSE by 0.29 grades, which is equivalent to 0.19 standard deviations in the population. Foundation tier shows a similar structure, and with school fixed effects included, a 1 standard deviation increase in the peer ability measure leads to a 0.2 grade increase at GCSE, which is equivalent to a 0.13 standard deviation increase in the population.

For mathematics, the coefficient is higher for the higher tier, but this decreases as we move through intermediate to foundation tier teaching. A one standard

[44] Standard deviation in higher tier English is 0.654 and foundation tier is 0.502.

deviation increase, within the tier, in the class average at key stage 3[45] leads to an increase in individual pupil's attainment of approximately 0.36 grades, or 0.2 standard deviations, in higher tier, 0.34 grades, or 0.2 standard deviations, in intermediate and 0.35 grades, or 0.2 standard deviations in foundation tier.

5.4.4 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 22 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. 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 0.11 and 0.13 grades, or between 0.07 and 0.08 standard deviations, for a one standard deviation increase in the class average measure. For the foundation tier, a significant positive effect is seen, equivalent to an increase of between 0.15 and 0.17 grades, or between 0.09 and 0.11 standard deviations, for a one standard deviation increase in the within tier class average measure.

[45] Standard deviation in higher tier mathematics is 0.468, intermediate tier is 0.518 and foundation tier is 0.879.

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 21 with higher estimates of peer effects in mathematics and especially higher tier mathematics and low estimates in English. However, the apparently random setting schools for English suggest there was a moderately large upward bias to the estimates in Table 21. This may be because of there only being two rather than three exam tiers in English.

Examining the effects for mathematics, for the higher tier entry, strongly significant positive effects are once more observed. The estimated coefficients indicate a 1 standard deviation increase in the class average measure, within the tier, is associated with a 0.38 grade, or a 0.21 standard deviation, increase in individual outcomes. For intermediate tier, again there are statistically significant positive effects of a more able peer group. The coefficients indicate that a 1 standard deviation increase in the within tier peer ability measure is associated with a 0.4 grade, or a 0.22 standard deviation increase in individuals outcomes. For foundation tier, the story is a little less clear cut, with a small positive effect observed in specification 4, although the statistical significance is greatly reduced in specification 5, with smaller effects observed than in the OLS case.

5.4.5 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 by Lefgren (2004b). Table 23 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 22 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 21 but very similar to the apparently random sample seen in Table 22.

The estimates for mathematics 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 22. 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 mathematics are insignificant from zero and significantly different from the estimates in Table 22. This alternative approach produces result very much in line with our apparently random sample except for higher tier mathematics.

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 least squares regression. Table 24 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 mathematics, 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 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 mathematics 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, or 0.42 standard deviations, by being in the higher tier classroom than in the lower tier classroom and for mathematics an average improvement of 0.62 grades at GCSE, which is equivalent to 0.34 standard deviations.

5.5 Conclusions

We find significant evidence of non-trivial 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.

For these schools with near-random allocation of pupils within the tier, some caution needs to be taken in trying to generalise these results to the population as the whole. Due to the complexity of the transmission mechanisms involved with the effect of a peer group, it is possible that different effects could be observed when rigidly grouping students according to ability than when they are randomly assigned to classrooms. For example, in a school with very rigid setting policies, teachers will be able to set the level of their classes according to the majority of the pupils. However, in randomly assigned classes, teachers will have to set the level of teaching differently, so this could have a further effect on the pupils.

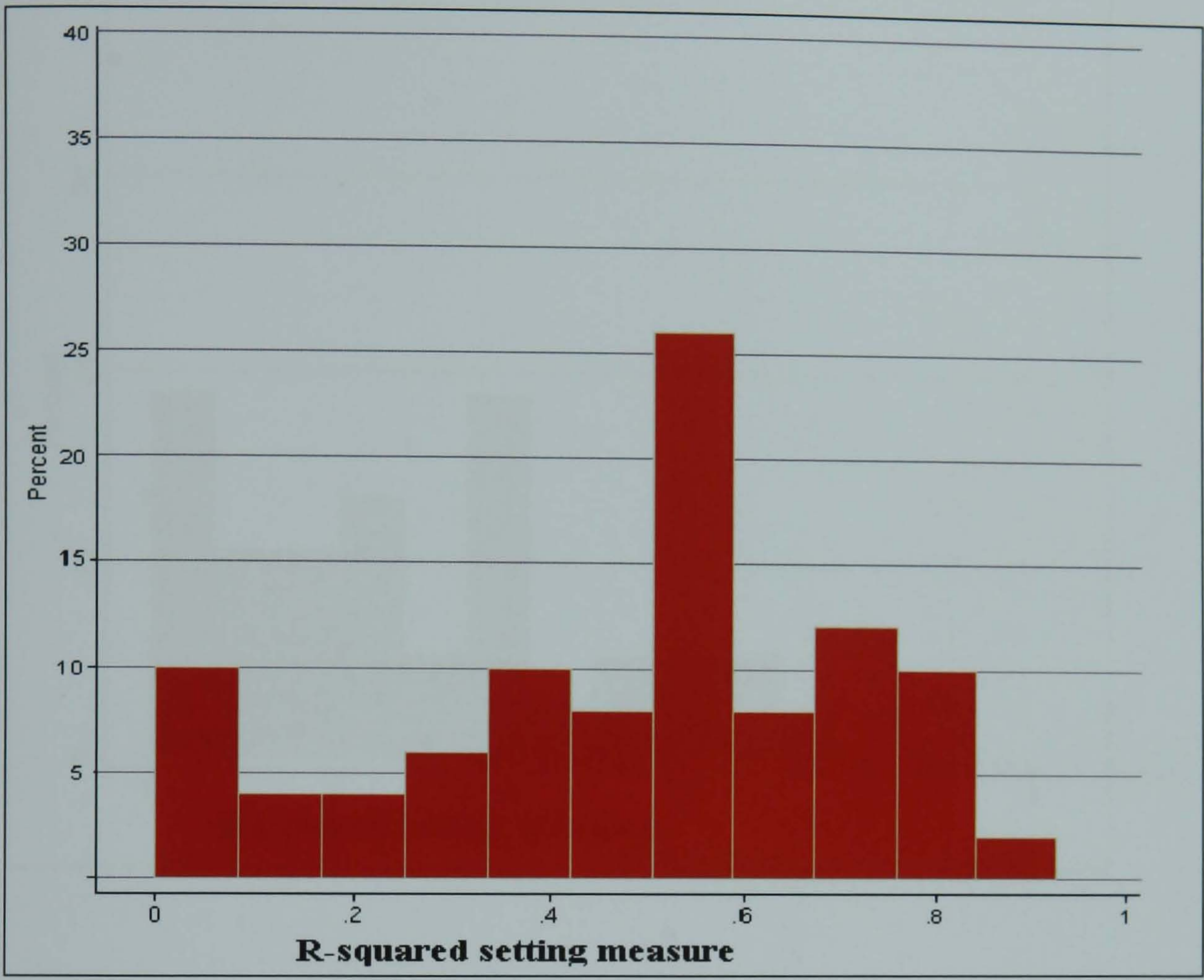
We find very similar results using the sub-sample of schools that apparently allocate pupils (near) randomly within an exam tier and for the two stage least squares approach. These estimates of the effect of a more able peer group are approximately one fifth to one half of the unconditional OLS specifications 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 two stage least squares estimates, except for within higher tier mathematics and English. It is apparent from Figure 11 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 two stage least squares specification. However, the two stage least squares 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 mathematics 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. It is suggested by Hallinan and Sorensen (1985) and Kubitschek and Hallinan (1998) that the process of tracking fosters friendships due to similarities between pupils. As such, a tracked group may foster closer friendships than an untracked group, and could increase the effect of a more able peer group.

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.

Figure 10 Distribution of R-squared setting measure for English between schools

a) Whole schools



b) Higher tier

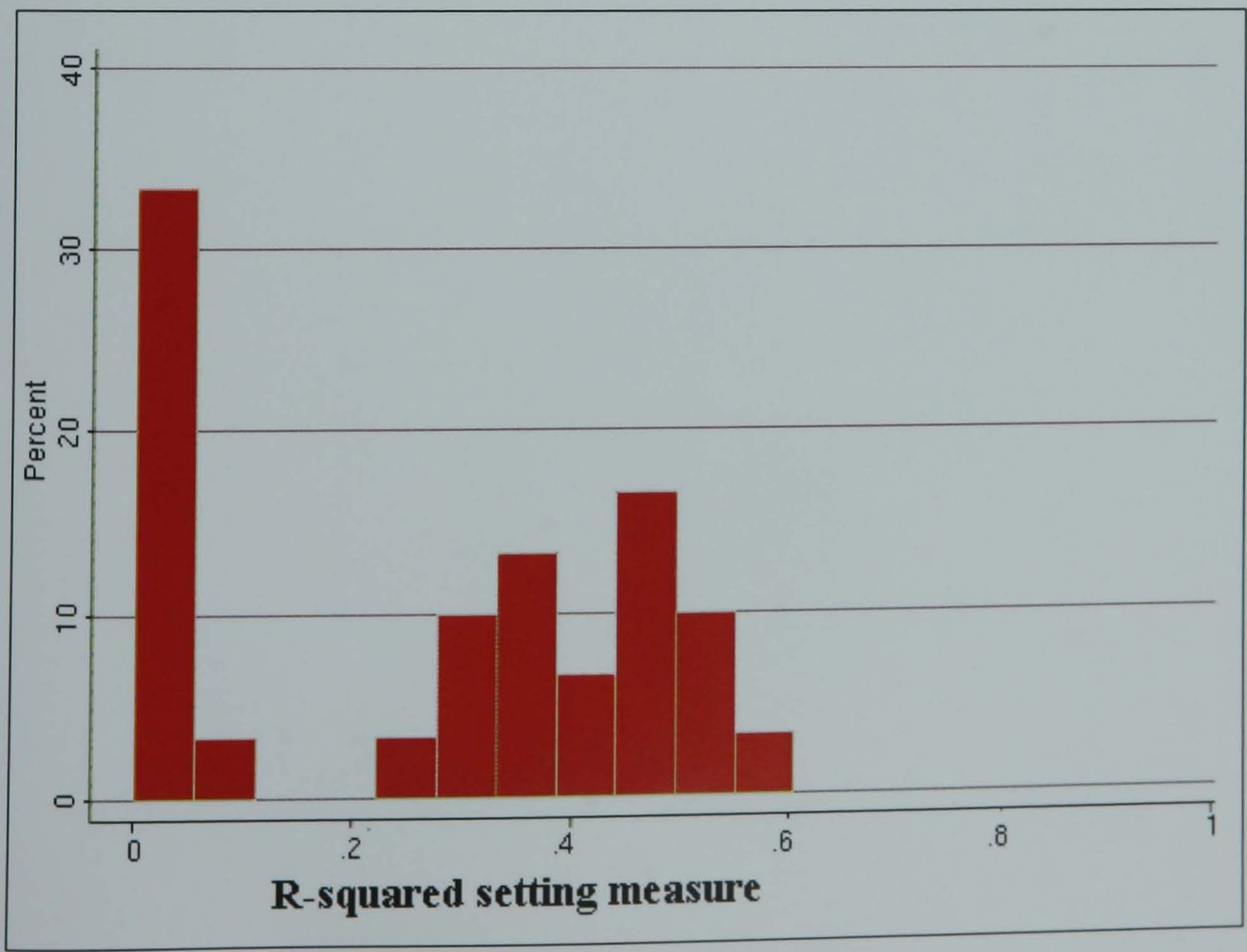


Figure 10 (Continued)

c) **Foundation tier**

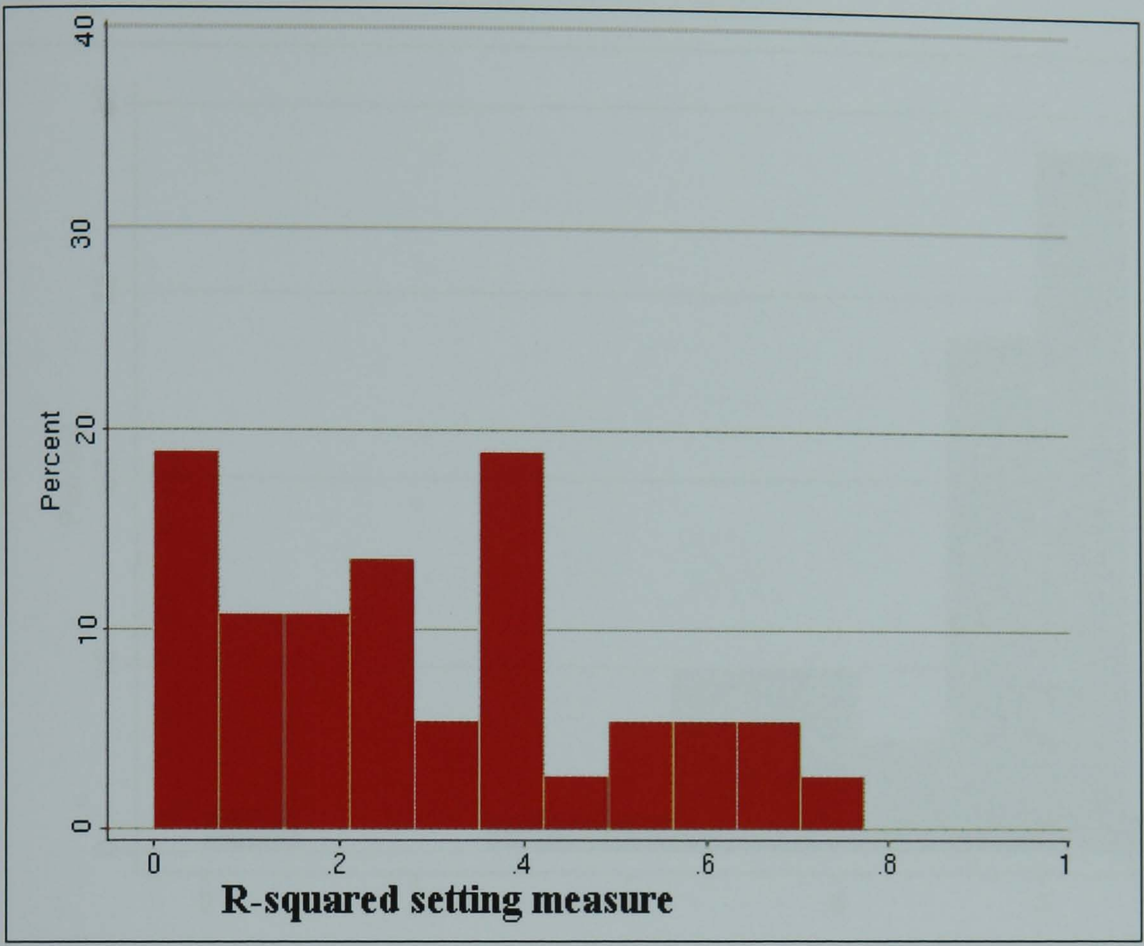
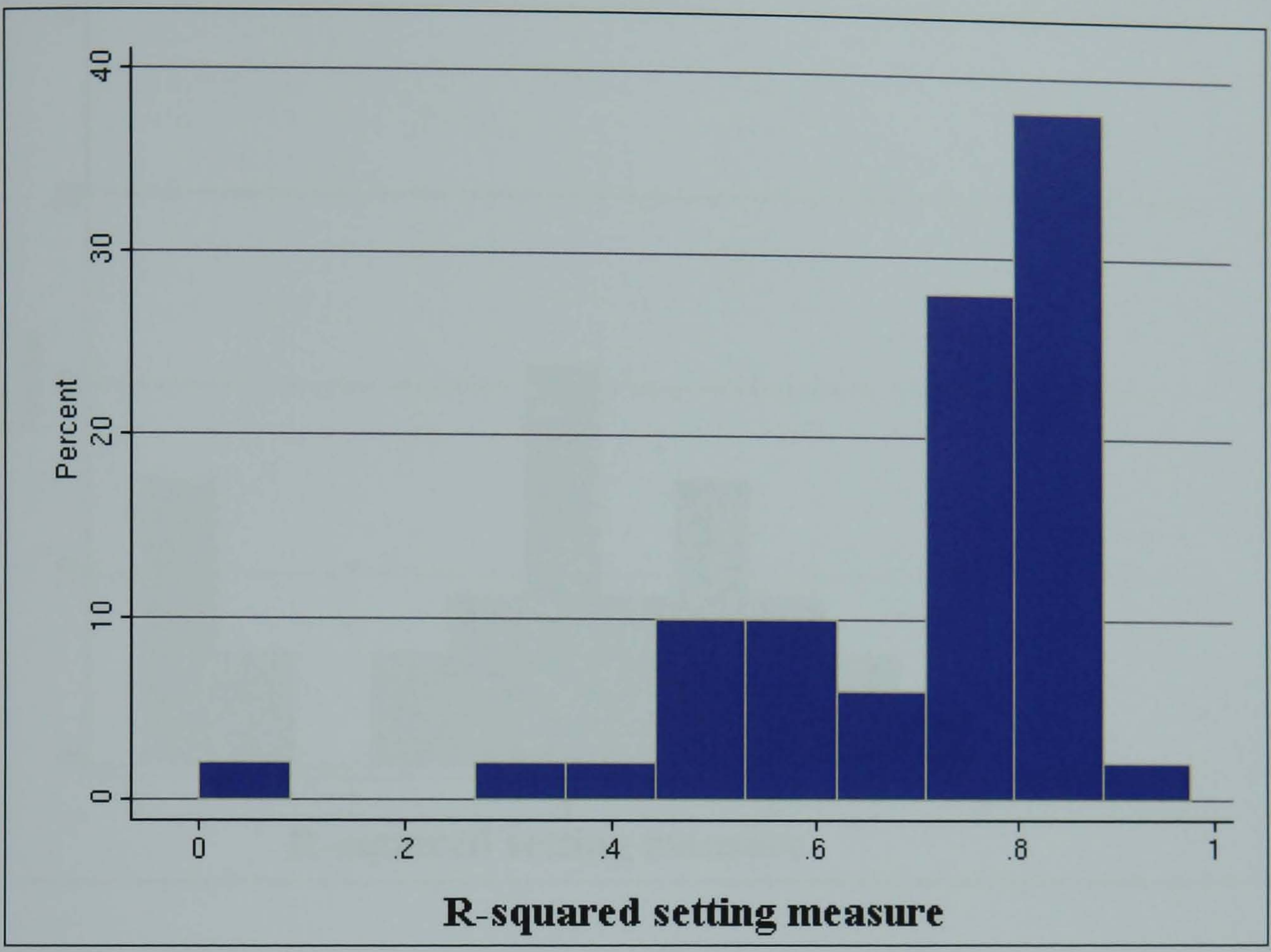


Figure 11 **Distribution of R-squared setting measure for mathematics between schools**

a) Whole schools



b) Higher tier

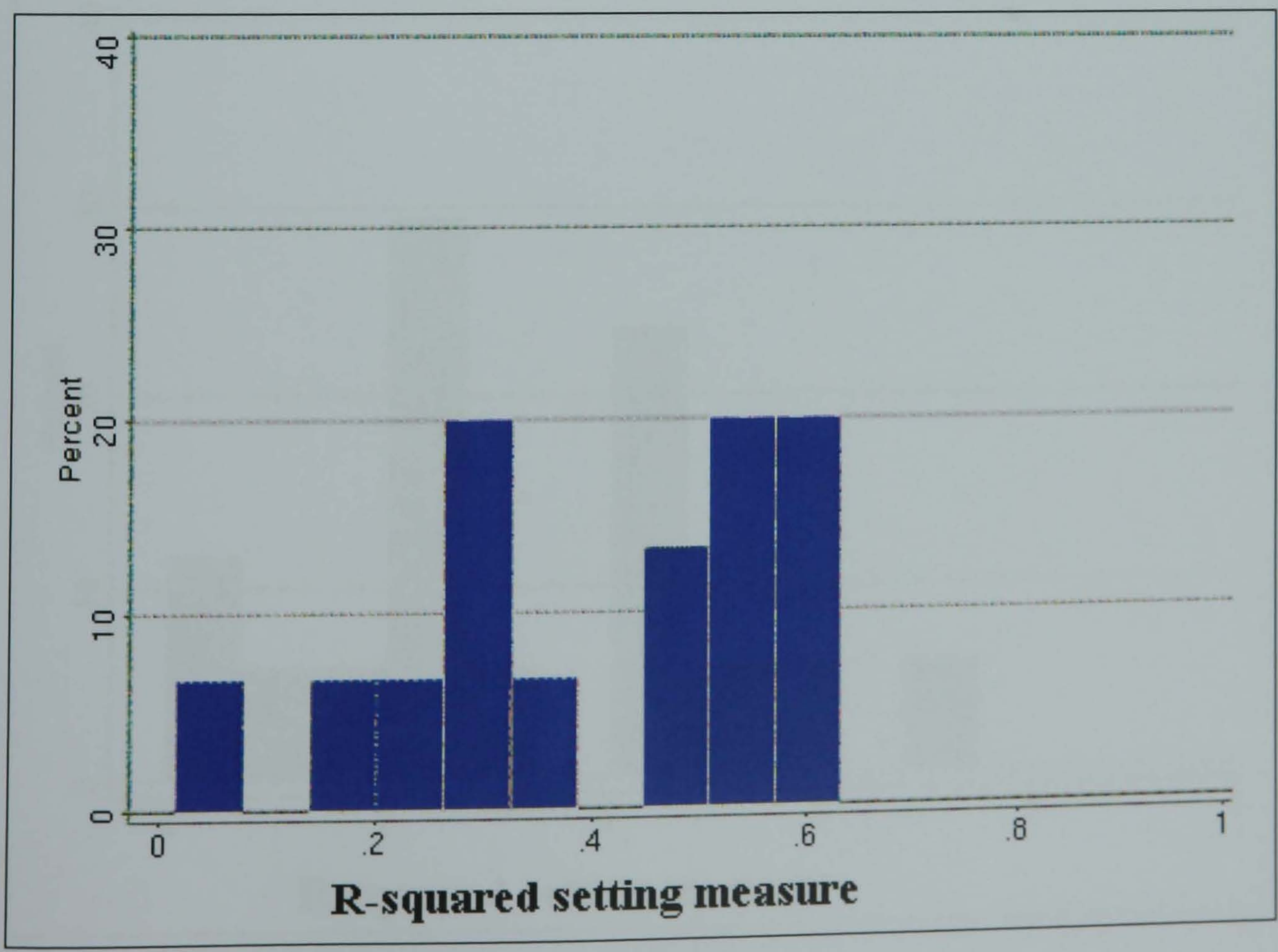
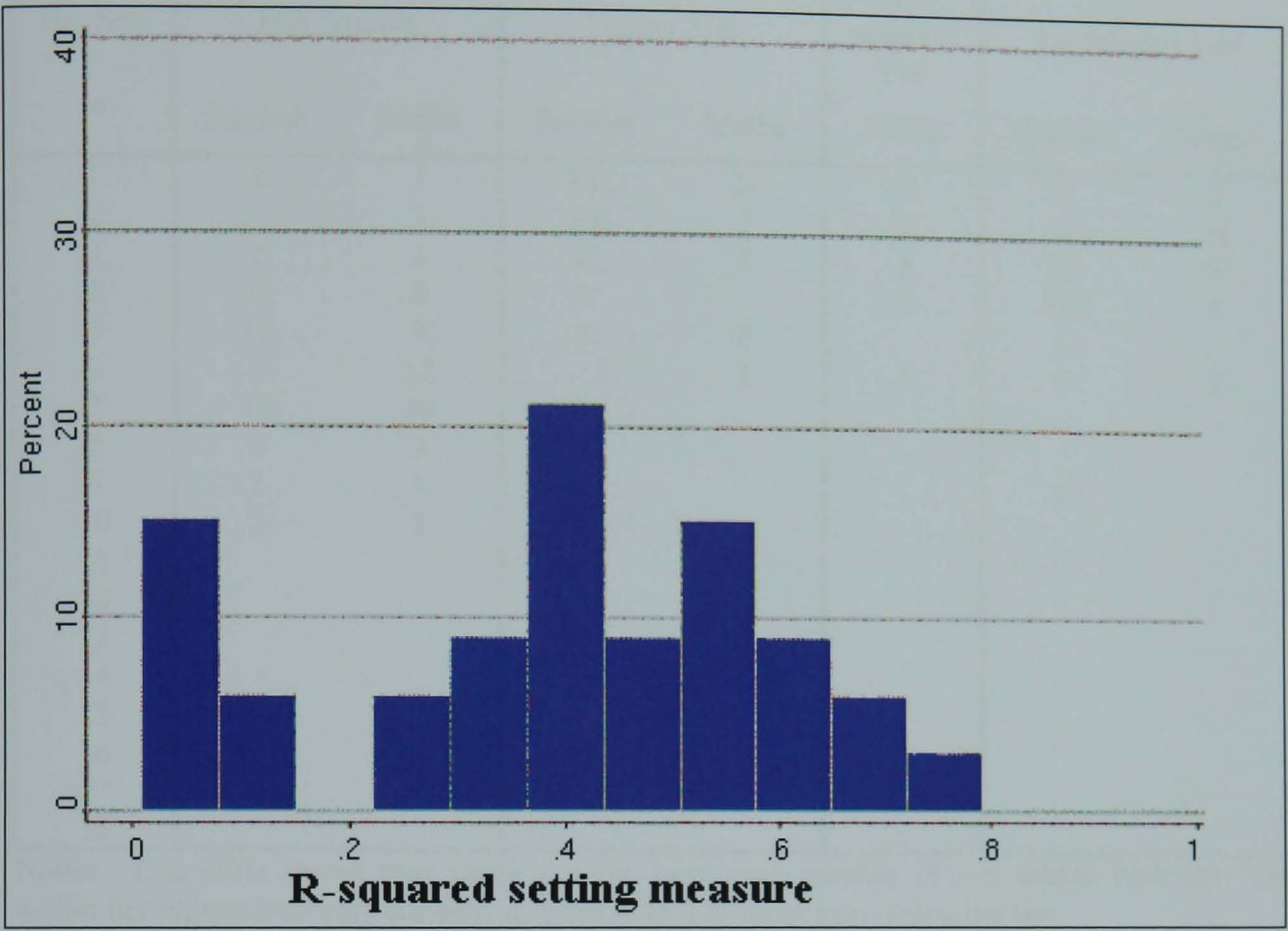


Figure 11 (continued)

c) Intermediate tier



d) Foundation tier

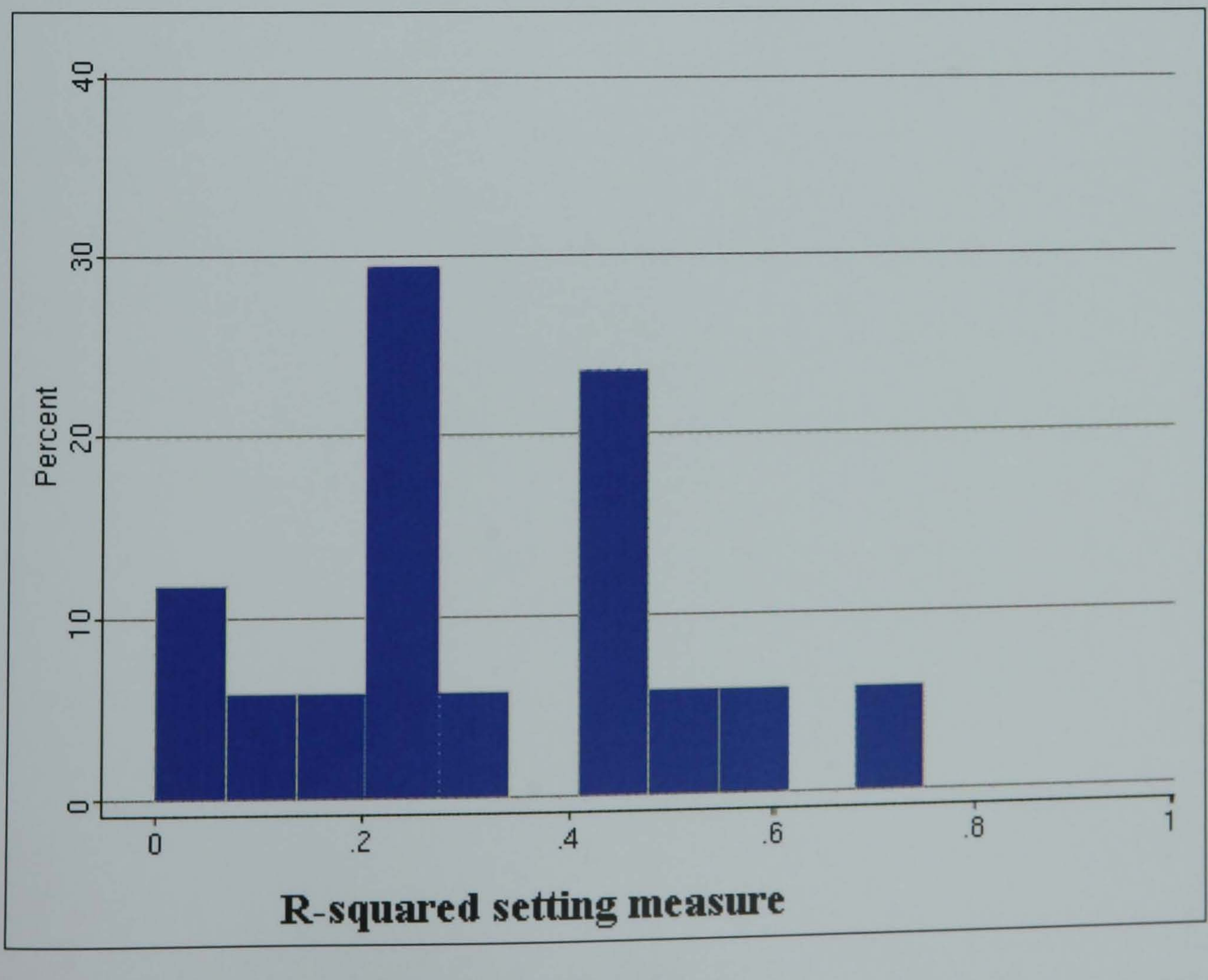


Table 18 Number of school/years with specified number of sets

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 19 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)	
Mathematics	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)	
Mathematics	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)	
Mathematics	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)	
Mathematics	6.742	6.942	16.258	0.350	0.605	446
	(0.834)	(0.697)	(0.292)	(0.477)	(0.032)	
Intermediate tier						
Mathematics	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)	
Mathematics	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 observation is an individual child. Low R-squared value is defined as being lower than 0.35. Large R-squared value is defined as being larger than 0.4

Table 20 Description of regression specifications

Specification	Description
1	Includes 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.
2	Includes the subject specific key stage 3 score
3	Includes the other subject key stage 3 scores
4	Includes school fixed effects
5	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)
6	Subsample of 4 with identifiers for teachers who teach 2 or more classes
7	Only with teachers who teach 2 or more classes. (Missings and teachers who teach only one class omitted)

Table 21 Results from ordinary least squares estimation of the effect of peer ability on outcomes

	1	2	3	4	5	6	7
School/year fixed effects				√	√	√	√
Teacher fixed effects					√		√
English							
Full Sample							
Class Average peer measure	1.169***	0.558***	0.336***	0.439***	0.442***	0.437***	0.425***
	(0.031)	(0.036)	(0.032)	(0.033)	(0.032)	(0.044)	(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***	0.412***	0.248***	0.442***	0.447***	0.761***	0.862***
	(0.071)	(0.073)	(0.065)	(0.066)	(0.070)	(0.174)	(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***	0.305***	0.224***	0.367***	0.435***	0.357***	0.238
	(0.064)	(0.068)	(0.063)	(0.055)	(0.063)	(0.123)	(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***	0.676***	0.555***	0.605***	0.595***	0.632***	0.613***
	(0.021)	(0.046)	(0.047)	(0.045)	(0.045)	(0.065)	(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***	0.699***	0.571***	0.758***	0.767***	0.884***	0.919***
	(0.088)	(0.117)	(0.113)	(0.079)	(0.070)	(0.145)	(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***	0.542***	0.441***	0.630***	0.650***	0.728**	1.055**
	(0.084)	(0.090)	(0.088)	(0.081)	(0.083)	(0.264)	(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***	0.502***	0.375***	0.457***	0.400***	0.926***	0.894***
	(0.085)	(0.092)	(0.092)	(0.070)	(0.065)	(0.273)	(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 20. All regressions include controls for 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. 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 22 **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		Mathematics	
	4	5	4	5
School/year fixed effects	√	√	√	√
Teacher fixed effects		√		√
1. Higher Tier				
Low R-squared				
Class Average peer measure	0.198**	0.167**	0.820***	0.806***
	(0.082)	(0.069)	(0.182)	(0.113)
Observations	1330	1330	469	469
R-squared	0.67	0.67	0.43	0.47
High R-squared				
Class Average peer measure	0.524***	0.518***	0.773***	0.792***
	(0.097)	(0.118)	(0.115)	(0.109)
Observations	770	770	701	701
R-squared	0.63	0.63	0.66	0.67
2. Intermediate Tier				
Low R-squared				
Class Average peer measure			0.772***	0.769***
			(0.139)	(0.184)
Observations			633	633
R-squared			0.47	0.48
High R-squared				
Class Average peer measure			0.626***	0.634***
			(0.109)	(0.112)
Observations			1144	1144
R-squared			0.42	0.43
3. Foundation tier				
Low R-squared				
Class Average peer measure	0.291***	0.331***	0.296**	0.230*
	(0.090)	(0.099)	(0.143)	(0.135)
Observations	936	936	588	588
R-squared	0.39	0.40	0.44	0.44
High R-squared				
Class Average peer measure	0.446***	0.545***	0.556***	0.512***
	(0.091)	(0.098)	(0.079)	(0.075)
Observations	556	556	933	933
R-squared	0.46	0.47	0.54	0.56

Notes Dependent variable is the GCSE score in English or mathematics. Specifications of regressions shown in Table 20. All regressions include controls for age, gender, prior achievement, 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. Method of estimation is ordinary least squares. (OLS) Robust standard errors, clustered at classroom level, are shown 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 23 Results from two stage least squares estimation of the effect of peer ability on outcomes

	English		Mathematics	
	4	5	4	5
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 20. All regressions include controls for age, gender, prior achievement, 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. Method of estimation is two stage least squares. Robust standard errors, clustered at classroom level, are shown in parentheses. * indicates significant at 10%; ** indicates significant at 5%; *** indicates significant at 1%.

Table 24 Test of endogeneity of class peer ability measure

	English		Mathematics	
	4	5	4	5
School/year fixed effects	√	√	√	√
Teacher fixed effects		√		√
Higher Tier				
Class Average peer measure	0.377***	0.380***	0.249	0.201
	(0.128)	(0.136)	(0.225)	(0.196)
Residuals	0.079	0.080	0.566**	0.627**
	(0.130)	(0.137)	(0.278)	(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***	0.671***
			(0.214)	(0.210)
Residuals			0.056	-0.024
			(0.217)	(0.217)
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)	(0.164)	(0.156)	(0.173)
Residuals	0.068	0.141	0.175	0.152
	(0.127)	(0.174)	(0.167)	(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 20. Method of estimation is ordinary least squares. (OLS) Robust standard errors, clustered at classroom level, are shown in parentheses. * indicates significant at 10%; ** indicates significant at 5%; *** indicates significant at 1%.

Chapter 6 Peer Effects in English Primary Schools.

An IV estimation of the effect of a more able peer group on age 11 examination results.

6.1 Introduction

In chapters 4 and 5, I have examined aspects of the effect of a child's peer group on their outcomes, and the difficulties that exist in estimating these effects. In chapter 5, I estimated the effect of a more able peer group in a small sub-sample of secondary schools. However, there is a worry that, since this was only a small sub-sample of schools, the estimated effects may not be representative of all schools. Further, in chapter 5, I have only estimated the effect of a more able peer group in secondary schools. This chapter again estimates the effect of a more able peer group on outcomes, this time in primary schools, taking advantage of a two stage least squares strategy to identify these peer effects.

There is strong evidence that pupils' outcomes in compulsory national assessments are strongly influenced by their month of birth, (Sharp (1995), Crawford et al (2007), Strom (2004)), with the oldest pupils within the year group performing better than their younger peers. In England, these differences are usually attributed to the older pupils gaining more maturity as they sit the examinations when they are older than the younger pupils. This correlation between individual outcomes and month of birth suggests also that a peer group that consists largely of older pupils will, in general, have higher previous outcomes.

This chapter uses an identification strategy which takes advantage of this correlation between the month of birth and outcomes in externally assessed examinations, with pupils born in September having an advantage over those born in August to carry out an instrumental variables analysis of the effect of a more able peer group on the outcomes of children at age 11. I take advantage of a within school estimation, conditioning on prior achievement. Furthermore, I suggest that the observed effects are credibly the effect of a more able peer group, rather than being confused with the effect of an individual having an older peer group. I contend that whilst it may be advantageous for children to be born in September, the proportion of pupils born in each third of the year is essentially random, and this is backed up by the Hansen J test of overidentifying restrictions, which suggests that the instrument is credibly exogenous in all specifications.

In order to examine the direct effect of a more able peer group within classrooms, I use the same characteristic of schools that I take advantage of in chapter 4. That is, I classify schools with 30 or fewer pupils within the cohort as schools with credibly only one class per cohort. As such, all of the pupils within the school-year can be defined as the peer group who are taught in direct contact with each individual pupil. Additionally, this chapter examines differential effects of a more able peer group on pupils who are close in terms of ability to the ability of their peer group and on pupils whose ability is a long way from the ability of their peers.

Testing of the validity of the instrument suggest that it is exogenous in all specifications, and I find significant, non-trivial, positive effects of having a more able peer group on results at key stage 2 in English and mathematics, with a larger effect being observed in mathematics than in English. Furthermore, the results suggest that in both English and mathematics the strongest effects of a more able peer group are observed for children who have prior outcomes that are close to the average outcome of their peer group, with a reduced effect observed for pupils who are a long way from the ability of their peer group. However, the effects look roughly symmetrical around the peer group, with only the pupils who have outcomes which are a long way above the ability of the peer group in English within small schools showing insignificant effects.

This chapter begins by examining the literature related to age within year and ability or outcomes, and also examines prior literature where age has been considered as an instrument within education. I then look at specific data issues faced from the PLASC dataset utilised in this chapter, and the specifics of the data required for the statistical analysis. Section 4 will examine the methodology used, whilst section 5 will discuss the results gained from the statistical analysis, and I will finish with conclusions based on the results and further discussion.

6.2 Literature

This chapter uses the age of the peer group as an instrument for the ability of the peer group, and in order for this instrument to be valid, it is necessary for there to be an appreciable correlation between student's outcomes and their age within the year. A full discussion of the peer effects literature is given in chapter 3.

Bell and Daniels (1990) consider whether children born in the summer are disadvantaged in education. They find that summer-born pupils perform worse than their autumn-born peers at ages 11, 13 and 15, with the effect decreasing as the children get older. However, since younger pupils could start school later in the academic year, there may also be a length of schooling effect, but they conclude that this is small compared to the overall age effect by examining foreign studies without the length of schooling effect. Borg and Falzon (1995) use Maltese data to investigate the role of pupil age in the birth date effect in scholastic achievement, and also consider the role of sex differences. They acknowledge that there is a problem in assessing the magnitude and direction of the effects of age and gender, as the composite scores used may have different age and gender effects in their constituent parts and so the effect may be subsumed or cancelled out. Their results find that the oldest pupils perform better in Maltese, English and mathematics than the younger pupils, and find that girls consistently outperformed boys in the three school subjects across three age groups.

Melkonian and Ierokipiotis (1997) use data from Cyprus to investigate the variations in educational achievement based on the position of the child's birth within the year. Their results contrast with most of the accepted literature as they find that the youngest students outperformed the oldest students. However, this is justified by the Cypriot education policy of promoting students into the following year based on their examination results, and as such, the youngest students in an age group made up, by far, the highest percentage of those students repeating a year.

Looking at English data, Sharp (1995) examines the effect of season of birth on outcomes at both key stage 1 and GCSE examinations, and finds that the eldest children within the schools perform best in these assessments. Likewise, Sharp et al (1994) consider the effect of season of birth on academic outcomes at age 7. They find that the oldest pupils perform best, but their analysis is clouded by some of the younger pupils only having eight terms of education, compared with their peers who have received nine terms of education.

Crawford et al (2007) analyse the impact of when a child is born on outcomes in English schools. They compare outcomes for children within schools who are born in September with those born in August, and control for other factors that are likely to affect children's outcomes. They find that August born boys and girls are at a significant disadvantage to their September born peers, but that this disadvantage decreases over time. They quantify that at age 5, August born boys are 0.817 standard deviations (SDs) behind September born boys, whilst August born girls are 0.768 SDs behind September born girls. However, by age 16, this has decreased to August born boys being 0.131 SDs behind their September born counterparts, whilst for girls; the penalty of being born in August is 0.116 SDs. Furthermore they examine pupils with special educational needs, including both children with statements of special educational needs and children with non-statemented special educational needs. They find that at age 11, August born girls are 25% more likely to have statements of special educational needs, whilst the boys are 14% more likely to have statements. However, this difference falls back at age 16. However, they argue that the identification of these special educational needs, particularly for those non-statemented children may simply be due to them progressing at a slower rate than their older peers. They argue that the major reason for the August-born penalty is that August born children are essentially a year younger than their September born counterparts when they sit the tests. These issues are now widely recognised amongst UK policy makers (see BBC (2008)). Following this review, the secretary of state responsible for education within the Department for Children, School and Families (DCSF) launched a review of primary education, and the minister suggested that summer born children should be allowed to defer their entry into school by up to a year.

Further weight to the argument that August-born children are likely to perform worse in academic testing is added by Strom (2004), which examines the effect of birth-date on children's outcomes in formal testing at age 15-16. He finds a significant disadvantage for the youngest children in reading compared with their older classmates.

6.2.1 Choice of instrument.

Atkinson et al (2006) use age within year as an instrument for whether a pupil attends a grammar school or not. They state that "Within year age has a direct effect on attainment at 16: in both selective and non-selective LEAs, older pupils achieve higher GCSE scores" (Atkinson et al (2006,25)). Angrist and Krueger (1991) use the age of a child within a school year based on the quarter of the year in which they are born as an instrument for education level. They find no significant difference from their OLS results. However, Angrist and Krueger's approach is criticised by Bound et al (1995) who demonstrate that Angrist and Krueger's results are strongly affected by including additional instruments in the analysis, and their results are subsequently biased due to their instruments being weak. Angrist and Krueger's choice of instrument tries to capture the length of time students spend in education, due to the fact in some US states; all children start school at the same time, but are allowed to leave school directly after their 16th birthday. This system is not reproduced in the UK, and as discussed above, there is an appreciable difference in outcomes associated with the birth-date of the child.

Sandgren and Strom (2005) examine explicitly the effect of an older peer group on educational outcomes. In order to estimate the effect of an older peer group, they take advantage of the Norwegian schooling system, which has similar characteristics to the English schooling system. Pupils are enrolled at the same time, and are educated without grade repetition, so the only effect on outcomes observed is an age effect, rather than a length of schooling effect. They examine the effect of an older peer group on outcomes, and find a significant effect in mathematics and reading, with the effect more robust for male students than for female students. However, they do not try to examine the effect of a more able peer group. This chapter deviates from Sandgren and Strom (2005), as rather than examining the effect of a change in the age make-up of the peer group, I use the age make-up of the classroom as an instrument for the ability of the peer group.

Maurin et al (2005) attempt to estimate the effect of a peer group on individuals' outcomes in English primary schools in 2002-2003. Their identification strategy is similar to the strategy I employ in this chapter, and uses the proportion of the peer group born in each month to try and identify their effects. However, like Sandgren and Strom, they estimate the effect of an older peer group on the outcomes of individuals rather than explicitly considering a two stage least squares methodology. Their results suggest the existence of non-linear peer effects, but they cannot disentangle the possible effect of having an older peer group from the effect of having a more able peer group.

6.3 Data

This chapter uses data from the Pupil Level Annual School Census (PLASC) and the National Pupil Database (NPD), as described in chapter 2. This chapter takes advantage of data up until 2006, so adds another two years of data from that used in chapter 4. The pupils in this dataset were examined in key stage 2 examinations between 2002 and 2006, and were examined in key stage 1 examinations between 1998 and 2002. These key stage 1 test scores are necessary to model the ability of the peer group. All single sex schools are removed from the data, as are any schools that appear fewer than 3 times in the dataset. Further, any school that has fewer than 10 pupils in the cohort is dropped from the sample.

At key stage 1, pupils are examined in reading, writing and numeracy. In order to create measures of English and mathematics, I consider the reading and writing as a composite English score, simply consisting of the average national curriculum level that the child achieved, and for mathematics, I simply take the numeracy score. These levels are subsequently normalized to mean zero and standard deviation of 1. However, it needs to be remembered that these key stage 1 scores are essentially discrete data, so an individual key stage 1 score can cover a relatively large range of abilities.

The scores used at key stage 2 are a much finer score, based on the raw score in the examination. However, these scores are not directly linked to a national curriculum level fixed across years. That is, the raw mark required in an examination to achieve a certain level one year is not necessarily the same score that is required in a subsequent year[46]. As such, this raw score is normalised by year to have the same mean and standard deviation as the national curriculum level score. As the national curriculum score is comparable across years, this can then normalised to have a mean of 0 and standard deviation of 1. This normalised raw score allows for a better comparison of outcomes of pupils than the very discrete and clustered measure of the national curriculum level achieved, where a pupil who achieves a good score within a level is classified the same as a pupil who achieves a borderline score in that level.

Whilst this strategy would also be desirable with the explanatory variable of key stage 1 achievement, the data is not currently available to consider a more continuous score. However, within the broad national curriculum level that the pupils achieve, there is also a smaller break-down into levels a, b and c, showing the pupils progression towards the next level.

The identification strategy I pursue requires that the peer test-score measure is correlated with the average age of the cohort, but that the average age of the cohort is uncorrelated with the error term. As such, it would be beneficial if the pupils are essentially randomly assigned to schools by age, and that parents do not try to maximise their children's outcomes by trying to ensure their children are the oldest within the academic year. For this identification strategy to be credible, we want there to be randomness on when children are born within the year, and so would expect an even spread of the month in which children are born (Although we would expect February to have significantly fewer births than October, since there are 28 (or 29) days in February, compared with 31 in October.

[46] For example, a mark of 43 at key stage 2 English in 2003 would have achieved a level 3, whilst the same mark in 2004 would have gained a level 4.

Figure 12 shows the distribution of the age of pupils at the start of the year when they take their key stage 2 examinations, based on month of birth. Key stage 2 examinations are sat in year 6, which is the academic year when pupils turn 11. There are small numbers of pupils who are in the wrong academic year based on their year of birth, but the vast majority of pupils are in a class with pupils born in the same academic year. In order to control for possible mis-codings of birth year, pupils who, in the raw data, are recorded as starting year 6 younger than age 9 or older than age 12 are dropped from the data. Whilst it might be hoped that the months of birth provide a perfect uniform distribution, this is not entirely the case, for several reasons. It is immediately apparent that there are fewer births in February than in any other month, but this is simply due to the fact that February is the shortest month[47]. There are also more than expected births in September[48]. However, when considering pupils born in the three quantiles of age based on months of birth shown in Figure 13 we can see that the distribution of births across the year is approximately equal. Figure 14 shows the distribution of the proportion of the within school cohort who are defined to lie in the youngest third of the age distribution, whilst Figure 15 shows the distribution of the proportion of the within school cohort who are defined to lie in the oldest third of the age distribution. These follow similar approximately normal distributions, as would be expected.

6.4 Methodology

I begin with a general educational production function, as discussed in chapter 3, considering pupils' attainment at key stage 2 to be a function of school inputs, consisting of school policy effects, teacher effects, and peer ability effects, family inputs and demographics affecting the ability of the child to learn effectively.

There are a large number of factors that are likely to be constant within a school, such as the neighbourhoods in which the children grow up, and there will be high

[47] We would expect the proportion of births in February to be 0.077 (2sf). The observed proportion we see here is 0.077 (2sf)

[48] The proportion within our sample who are born in September is 0.087, compared with a theoretically random proportion of 0.082.

correlation between pupils in the level of parental income within neighbourhoods and schools. Furthermore, all of the pupils within the schools are taught in the same atmosphere, with the same facilities available to them, with the same teaching culture that is engendered by the head teacher. As such, it is necessary to try to control for these factors, which may also be correlated with the ability of the peer group, but which are not directly observable within my data. In order to control for these correlated between school heterogeneities, I use similar techniques to Hanushek et al (2003) and McEwan (2003), and include school fixed effects.

As such, I model attainment A at time t for individual i in school k to be a function of prior attainment, individual family inputs, F, within school cohort effects, S, which includes an underlying school effect and the effect of a more able peer group, the underlying ability of the pupil μ and error terms, ε.

In order to control for inputs in prior periods which are not observed, I include lagged academic outcomes as a measure of the cumulative of these previous inputs up to key stage 1 examined at age 7, in order to consider the effect of the presence of a more able peer group between the ages of 7 and 11. Pupils' prior achievement can thus be modelled from their key stage 1 examinations, sat at the age of 7, that is, 4 years previously.

$$A_{t,ik} = f(A_{t-4,ik}, F_{t,ik}, S_{kt}, \mu_{ik}, \varepsilon_{t,ik}) \qquad (16)$$

This chapter considers the effect of being in a school with a more able peer group. There is still the worry that, despite having included school fixed effects, there are elements within the error term that are correlated with both the outcome at age 11 and the prior ability of the peer group. In order to mitigate this correlation, and the resultant bias inherent I use a similar strategy to Sandgren and Strom (2005) and Maurin et al (2005). I use two stage least squares to estimate the effect of a more able peer group. I use the proportion of pupils within the school-year who lie in the oldest third and the proportion of pupils within the school-year who are in the youngest third of the age distribution as instruments for the average within year school average score at 7.

The first stage of the two stage least squares is estimated thus:

$$classave_{j\neq i,k,t-4} = a + ageave_{k,j\neq i}\beta_0 + X_{ikt}\beta_1 + KS1_{ikt-4} + s_k + t_t + u_{it} \quad (17)$$

where *classave* is the average key stage 1 score of the pupils in English (or mathematics) gained in school k, by all of the pupils j at time t-4. *ageave* is a vector containing the proportion of pupils within the school cohort that are in the top third of the age distribution and the proportion of pupils who lie within the bottom third of the age distribution. s is a school level fixed effect and t is a dummy for the year that the students sit the key stage 2 examination and u is a random error term. There is little variation within the school cohort of the *classave* variable, as the only variation comes from the omission of individual pupils.

Since we expect there to be correlations between the explanatory variables at a school level, due to factors explained above, it is necessary to adjust the standard errors to mitigate the problems when the independent and identically distributed assumptions are dropped. In order to control for these effects, I cluster the standard errors at school level.

As discussed in the section 6.2, we would expect there to be a correlation between the ages of pupils within the cohort with their individual outcomes. That is, we would expect the oldest pupils to gain the highest grades at key stage 1. Therefore, we would also expect a cohort with a high proportion of 'old' pupils to have a better average outcome than a cohort with a high proportion of 'young' pupils. As such, we would expect the proposed instruments to be strongly correlated with the ability of the peer group. Since the standard errors are clustered at the school level, this implies that the observations are no longer independent and identically distributed, and as such, I need to appeal to the methods proposed by Kleibergen and Paap (2006) in order to test for underidentification of the endogenous variables. (see Baum et al (2007)). As such, I use the Lagrange multiplier (LM) test proposed by Kleibergen and Paap (2006) to test for underidentification. Since the sample is close to a population.

there is little worry that if we reject the null of underidentification that there will be a problem with weak instruments. However, I do calculate the Kleibergen Paap F-statistics and compare them with the critical values calculated by Stock and Yogo (2005), but do not report them here.

I estimate the second stage of the two stage least squares thus

$$KS2_{ikt} = \alpha + classave_{k,j\neq i,t-4}\gamma_0 + X_{it}\gamma_1 + KS1_{it-4}\gamma_2 + s_k + t_t + \varepsilon_{it} \qquad (18)$$

Where *KS2* is the individual pupil's *(i)* score at key stage 2 in English or mathematics. *X* is a vector of individual level characteristics, including pupil age, gender at time *t* and exam scores at time *t-4* (pupils take their key stage 2 examinations 4 years after their key stage 1 examinations) in English and mathematics at key stage 1.

6.4.1 Validity of the instrument

For the choice of instrument to be valid, it requires the month of birth to be credibly exogenous. Whilst there is a danger that some parents may try and influence the date of birth of their child, (See, for example BBC (2009)), it must be remembered that whilst they may have a preference, this is countered by difficulties in conception and by unintended pregnancies. Ford et al (2000) suggest that 28.7% of pregnancies in the Avon area are unintended, whilst Scheike and Jensen (1997) suggest that 59% of planned pregnancies take longer than 1 month to achieve conception. Further to these difficulties in achieving pregnancies, the time of birth is also difficult for parents to control, with 8.6% of UK births registered as premature, and only 33.7% of births occurring at the expected 40 weeks[49]. These factors suggest that the distribution of births will be credibly random. In order to further test these exogeneity assumptions, I appeal to statistical testing. Standard testing methods would appeal to the Sargan statistic. However, Baum et al (2007) suggest that since I consider the

[49] Premature birth indicates born before 37 weeks gestation. Source of statistics HES online, available at http://www.hesonline.nhs.uk/Ease/servlet/AttachmentRetriever?site_id=1937&file_name=d:\cfm files\1937\Accessing\DataTables\Maternity\Tables%2021%20to%2030\Tb27\Mat_Tb27_0708.xls&short_name=Mat_Tb27_0708.xls&u_id=8441 accessed on 15/09/09

possibility that there is correlation within clusters at school level, the Hansen J statistic is the correct statistic to consider when examining the test of overidentifying restrictions. That is, whether the instruments are truly exogenous.

6.4.2 Small schools versus large schools

In order to accurately assess the effect of a more able peer group, it may not be sufficient to simply examine entire schools. This is due to the fact that in large schools, there may be no interaction between pupils in different classrooms within the school. In order to try to observe pupils who are taught together, I use the same strategy as that used in Chapter 4. Since the PLASC data does not include classroom level data, I need to try to infer where pupils are directly taught with their entire school cohort. As such, in order to infer these classrooms, I consider schools that only contain 30 or fewer pupils in every cohort, which indicates that each year, the school only fills up one classroom, and then closes admissions.

6.4.3 Differential effects on different ability of pupils.

Previous studies have looked at differential effects of a more able peer group on outcomes for high ability and low ability children, for example Zimmer and Toma (2000). In order to examine this possibility, I consider the effect on individuals whose key stage 1 results are either close to the average outcome for their peer group, or for individuals who are far away from the mean ability of their peer group. To do this, I construct a new variable that measures the distance away from the peer group ability that an individual is, and construct quartiles of this distance on all of the individuals in the data

6.5 Summary statistics

Table 25 shows summary statistics for the key stage 2 outcome variables, the prior attainment at key stage 1, the peer ability measure (as measured by the average of the peers' scores at age 7), and the proportion of pupils within the peer group who lie in the oldest and youngest thirds of the age distribution within the school. The summary statistics are broken down into overall. between and within standard deviations.

As discussed within the data section, the key stage 1 and key stage 2 scores are normalised with mean 0 and standard deviation of 1. For this analysis. it is important that there is a significant variation within school of the peer ability measure. If not, then the peer ability will simply be absorbed by the inclusion of a school-level fixed effect. The distribution of the peer ability measure is shown in sections (a) and (b) of Table 25. Since the key stage 1 scores are normalised, then the mean of the peer scores is also zero. The standard deviation of the peer ability measure for English overall is 0.365 and the standard deviation of the peer ability measure for mathematics overall is 0.371. What is key is whether there is any variation within schools of this peer ability measure. For English, the standard deviation within school is 0.182, which makes up 24.8% of the total variance of the peer ability measure, whilst for mathematics, the standard deviation is 0.245, which makes up 43.5% of the total variance of the peer ability measure. As such, whilst the majority of the variation in the peer ability score is between school (56.6%-75.5%), there is still a significant within school variation in the ability of the peer group.

Similarly, in order for my instrument to be useful, there needs to be within school variation of the proportion of pupils who are in the oldest third of the age distribution and the proportion of pupils who are in the youngest third of the age distribution. The distribution of the proportion of pupils within the age thirds is shown in sections (c) and (d) of Table 25. Whilst the overall standard deviation is low, there is a range of proportions of pupils between 0 and 1 within the oldest third, and a range of pupils between 0 and 0.9 in the youngest third. Considering the within and between school variance in the proportion of the peer group in the youngest third of the age distribution, 80% of the variance is within the schools. Similarly, for the proportion of the peer group who are in the oldest third, 80% of the overall variance is within schools for the oldest third of the cohort. Furthermore, sections (e)-(h) of Table 25 suggest that 88% of the variance in the key stages 1 and 2 scores are within schools, which is as would be expected.

6.6 Results

In this section I will discuss the OLS and IV results for English and mathematics, initially examining the simple all school specification. I will then look at the effects for schools with a large distribution of examination scores, and examine differential effects of being in a class with a more able peer group for a student close to the average ability of the peer group and for one far away from the ability of the peer group.

I am interested in the estimates of the coefficients from equation (18):

$$KS2_{ikt} = \alpha + classave_{k,j\neq i,t-4}\gamma_0 + X_{it}\gamma_1 + KS1_{it-4}\gamma_2 + s_k + t_t + \varepsilon_{it}$$

I am particularly interested in the coefficient γ_0. In order to correct for possible endogeneity of the peer ability measure (*classave*), it is also necessary to consider a two stage least squares estimation using equation (17)

$$classave_{j\neq i,k,t-4} = a + ageave_{k,j\neq i}\beta_0 + X_{ikt}\beta_1 + KS1_{ikt-4} + s_k + t_t + u_{it}$$

Results of the first stage regressions are only reported for the most general specification.

6.6.1 OLS results

Table 26 gives OLS estimates of the effect of a more able peer group on outcomes in English and mathematics at key stage 2. In examining these results, I will begin by describing the estimates of the effects from the other explanatory variables, that is the variables which we do not suspect are endogenous, and will then move on to the estimates of the effect of a more able peer group.

As would be expected, we see a significant positive effect of own prior achievement on outcomes at key stage 2. For English, specification (ii) implies that a one standard deviation increase in prior achievement is associated with a 0.602 standard deviation increase in their key stage 2 scores, whilst a 1 standard deviation increase in mathematics scores at key stage 1 is associated with a 0.182 standard deviation increase in their English key stage 2 score. Furthermore, we can observe a strong negative effect of poor socioeconomic status, modelled by whether the child is eligible for free school meals. Also, in English, *ceteris paribus*, being male lowers the outcome at key stage 2 by 0.15 standard deviations. The only coefficient that isn't in the direction that might be expected is on the age of the child within the year. The direction of this coefficient can be explained by the fact that I have controlled for prior attainment. This is explained in Crawford et al (2007) that the gap between the oldest and the youngest decreases as the children get older[50]. Considering the effects of variables that we consider to be exogenous in mathematics, a similar set of effects are observed. The magnitude of the negative effect of free school meals is the same in mathematics, although, there is a stronger negative effect of an older pupil in mathematics. The largest difference is in whether the pupil is male or not. Having controlled for prior ability and age, boys perform 0.188 standard deviations better than girls in mathematics. However, this is as would be expected. As discussed in chapter 4, boys initially perform better in mathematics than girls, but this advantage is eroded over time. Finally, the prior achievement in English has marginally more effect on pupils' achievement at key stage 2 in mathematics than the prior attainment in mathematics had on scores at key stage 2 for English.

In terms of the effect of a more able peer group, Table 27 suggests that for both English and mathematics, a more able peer group leads to a reduction of the outcome at key stage 2, with a magnitude of a 1 standard deviation increase in the peer group outcome leading to approximately a 0.1 standard deviation decrease in the key stage 2 outcome score. However, it must be remembered that these estimates are likely to be correlated with the error term, and as such are likely to be a mis-estimate.

[50] As a robustness check it is possible to consider the specification both with prior achievement controlled for, and without prior achievement controlled for. The introduction of the prior achievement switches the direction of the coefficient on age from positive to negative.

6.6.2 Two stage least squares results

Table 27 gives results from the first stage of the two stage least squares estimation of the effect of a more able peer group on outcomes at age 11 for all pupils within all schools. In examining these results, I will begin by discussing whether the instruments I have used are plausibly valid based on econometric testing, and will conclude by discussing the effects of a more able peer group. As would be expected, the estimates of the effects of the variables that we do not suspect are endogenous are largely the same as those in the OLS estimation case.

The results presented here suggest a statistically significant negative correlation between the proportion of pupils who are young within the cohort and the peer ability measure, and a statistically significant positive correlation between the proportion of pupils who are old within the cohort and the average ability of the peer group. As such we would expect to reject the null of underidentification. Table 28 gives the results from the second stage of the two stage least squares for all pupils in all schools. For both English and mathematics, we can observe a significant and non-trivial positive effect of a more able peer group on outcomes at age 11 Reported in Table 28 are tests on the validity of the instruments under these specifications. As expected, based on the results from the first stage regressions, the P values on the Kleibergen-Paap test of underidentification are 0.0000 for both specifications (1) and (2) for both English and mathematics, and so we reject the null of no correlation between the instrument and the peer ability measure. The presence of a large sample and the size of the Kleibergen Paap LM test suggest that weak instruments should not be a problem. However, to check this, I compare the Kleibergen Paap Wald F statistic with the 10% maximal IV size statistic from Stock and Yogo (2005).[51] Table 28 also reports the Hansen-J test of the overidentifying restrictions. In both specifications, for both English and mathematics, we fail to reject the null that the instruments are not correlated with the error term. Since we reject the null of underidentification, and fail to reject the null of endogeneity of the instruments, our instruments appear to be valid.

[51] The Kleibergen-Paap F-statistics are calculated and compared with the Stock-Yogo critical values, and we reject the null of weak instruments at all reasonable levels of significance. These statistics are greater than the 10% maximal IV size in all samples.

Examining specification 1 and 2 for English in Table 28, the results imply that a 1 standard increase in the ability of the peer group leads to a 0.0423 standard deviation increase in the outcomes that a child achieves at key stage 2. Similarly, an increase in the peer group outcome in mathematics by one standard deviation is associated with an increase in a pupil's key stage 2 in mathematics score of between 0.0516 and 0.0597 standard deviations.

6.6.3 Results in small schools

As discussed in the methodology section, whilst we would like to observe directly the effect of a more able peer group within the classroom, this if often not possible, as for a large proportion of schools, we cannot directly observe which pupils are taught in a classroom with which. In order to estimate the classroom level effect, I consider here schools which have fewer than 30 pupils in all observed cohorts as a proxy for schools which teach all of their pupils in one class (which I describe here as a small primary school). Table 29 shows the results for all pupils who are educated within small primary schools. By placing the restriction on the size of the school, I have removed 8,863 schools from the sample, but we are still left with a large sample of children within the population (326,654 children in 3,056 schools).

Again, it is important to check the validity of the instruments. As with the all school sample, we strongly reject the null of no correlation of the instruments with the endogenous variables, and we also strongly fail to reject the null that the instruments are not correlated with the error term, so the tests support the argument that the instruments are valid.

The estimates of the effect of the exogenous variables are of the same magnitude as those observed in the full sample case, as is the estimate of the effect of a more able peer group. However, in contrast with the full sample case, there is only a significant positive effect of a more able peer group in specification (ii) for mathematics, and there is no significant effect within small schools of a more able peer group.

6.6.4 Results for pupils based on distance from the average of the peer ability

It is interesting to see if all pupils are affected in the same way by an increase in the ability of the peer group. That is, whether children who are a long way above the ability of their peer group would benefit as much from an increase in the ability of their peer group as children who are close to the ability of the peer group. For analysis here, I consider 4 quartiles of the distance of an individual pupils' key stage 1 score and the average key stage 1 score of their peers. Specification (a) is the lowest quartile below the ability of the peer group, (b) is the second quartile, (c) the third, and (d) is the highest quartile above the average outcome of the peer group.

Table 30 shows the results from all schools for English of the effect of a more able peer group on sub-groups of the population. Again, for all specifications, the tests for validity of the instruments reject the null of underidentification, and fail to reject the null in the test of overidentifying restrictions, indicating that the instruments are valid. Examining the coefficients on the effect of a more able peer group suggests that pupils who are closer to the ability of the peer group are affected more by an increase in the prior outcomes of their peer group than those who are a long way away. For specification (a), a 1 standard deviation increase in the average outcome of the peer group is associated with between a 0.115 and 0.119 standard deviation increase in a pupil's outcomes at key stage 2. Similarly, specification (b) suggests a 1 standard deviation increase in the peer ability leads to a between 0.154 and 0.165 standard deviation increase in the individual's outcomes. Specification (c) suggests a between 0.194 and 0.196 standard deviation increase, whilst specification (d) suggests between a 0.069 and 0.074 standard deviation increase from a 1 standard deviation increase in the peer ability measure.

Table 31 shows the effect of a more able peer group, broken down by distance from the peer ability outcome. Again, tests on the instruments indicate that there is no problem with underidentification, nor with endogeneity of the instruments.

The effects are of similar magnitudes to those seen in all schools, but the major difference is that in small schools, it appears that the most able students (i.e. the students whose ability is highest above the average ability of their peer group) do not gain any statistically significant advantage from being educated with a more able peer group.

Table 32 shows the estimates for a more able peer group in mathematics, again broken down by the distance of the individual pupil from the average ability of their peer group. The Kleibergen Paap and Hansen-J tests again do not find any problems with the instruments, indicating that the instruments are not invalid. The effect of other, exogenous, variables is of the same magnitude as that seen in the whole school regressions, other than for specification (b). Here, it appears that prior ability has no effect. Furthermore, these results suggest that an increase in peer ability will have a considerably larger effect on your own outcomes than for any other group. Comparing with sub-sample (a), for whom a 1 standard deviation increase in the peer ability measure leads to a between 0.167 and 0.181 standard deviation increase in key stage 2 score, for sub-sample (b) a 1 standard deviation increase in the peer ability measure is associated with between a 0.453 and a 0.459 standard deviation increase in the outcomes at key stage 2 in mathematics. As with English, sub-samples (c) and (d) see a reduction in the effect of a more able peer group on individuals' outcomes at key stage 2.
Table 33 shows the results within small schools, and shows a similar structure of effects, with the largest effects of a more able peer group once again seen for children who are close to the ability of the peer group, albeit below (i.e., sub-sample (b)). As with English, the significance of the effect of a more able peer group is reduced for sub-sample (d): that is, pupils whose outcomes at key stage 1 mathematics are a long way above those of the peer group.

6.6.5 Summary

In all of the specifications, I have rejected the null of underidentification and failing to reject the null of the excluded instruments being exogenous. These tests send a strong signal that the instruments are valid, and that there will be less bias from the IV estimates than from the OLS estimates. The IV estimates of the effect of a more able peer group suggest that an increase in the ability of the peer group by one standard deviation leads to an increase in the outcomes at key stage 2 by between 0.04 and 0.4 standard deviations. There is little difference between the estimates obtained within small schools and schools overall. However, it is clear that the strongest effect is observed for pupils who are close to the ability of their peer group.

6.7 Conclusions

In this chapter, I have examined the effects of a more able peer group on individuals' outcomes at age 11, with a sample of both full schools and a subset of pupils within the school based on how far the child is from the ability of the peer group. I have taken advantage of an instrument proposed by Angrist and Krueger (1991) as the age make-up of the peer group as an instrument for their ability. Whilst Sandgren and Strom (2005) suggest that there may be more mechanisms than just ability in operation when considering the effect of an older peer group on outcomes, my results show no evidence of any endogeneity of the instruments used. The results presented here suggest significant and non-trivial positive effects of a more able peer group on individual children's outcomes at age 11.

Estimates from the instrumental variables specifications suggest that a 1 standard deviation increase in the prior achievement of the peer group is associated with a between 0.04 and 0.4 standard deviation increase in the outcome the individual achieves at key stage 2. Furthermore, the results presented suggest that pupils who are close to the ability of their peer group benefit more from an increase in the ability of the peer group than those whose ability is further away from the ability of the peer group. Also, the results here imply that pupils who are a long way below the ability of their peer group are improved more by an increase in the peer group ability than those who are a long way above the ability of the peer group. This result is similar for the highest and lowest ability pupils to that presented by Zimmer and Toma (2000), who suggest that there is a greater effect of a higher ability peer group on lower ability pupils than for higher ability pupils, but this is in contradiction to the effect on high achievers observed by

Gibbons and Telhaj (2008) who suggest that there is a positive effect from a more able peer group on the highest and middle ability, but those at the bottom of the ability distribution are largely unaffected by an increase in the ability of the peer group.

In chapters 4, 5, and 6, I have estimated the effect of different characteristics of a pupil's peer group on their individual outcomes at ages 7, 11, 14 and 16 using various methodologies. In chapter 4, I estimated the effect of a change in the gender make-up on both boys and girls outcomes. The results obtained varied according to gender and subject. In English, a 10 percentage point increase (Approximately a 1 standard deviation change) in the proportion of the peer group that is female is associated with a between 0.01 and 0.03 standard deviation fall in boys outcomes, whilst a 10 percentage point increase in the proportion of the peer group that is female is associated with an approximately 0.01 standard deviation increase in outcomes for boys and girls in mathematics and science in primary schools.

It is interesting to compare the effects of a more female peer group with the effects of a more able peer group. In chapter 5, I estimated the effects of a more able peer group within the classroom at age 16. Estimates suggested that for English, within tier, a 1 standard deviation increase in the peer outcome measure is associated with a between 0.07 and 0.11 standard deviations increase in individuals' outcomes. For mathematics, chapter 5 suggested a 1 standard deviation increase in the ability of the peer group at age 16 was related to approximately a 0.2 standard deviation increase in individuals' outcomes. In chapter 6, the specification with all pupils suggests a 1 standard deviation increase in the ability of the peer group is associated with a between 0.04 and 0.06 standard deviation increase in individuals' outcomes in English and mathematics. These results are comparable in magnitude with those estimated in secondary schools for English, although they are smaller than those seen in mathematics. All of these results suggest that the effect of a more able peer group has more of an effect than for a more female peer group.

Comparing the results based on the ability gap between the pupil and their peer group, larger effects are observed. For pupils who are a long way above the ability of the peer group, a similar magnitude of effect is observed, with a 1 standard deviation increase in the peer ability being associated with a 0.07 standard deviation increase in individuals' outcomes. For the other three specifications for English, a one standard deviation increase in the ability of the peer group is associated with a between 0.1 and 0.2 standard deviation increase in the individuals' outcomes. For mathematics, the story is slightly different, with pupils whose results are close to the ability of the peer group, but below experiencing a 0.4 standard deviation increase in individuals' outcomes, whilst all the other effects are of a similar magnitude to those observed in English. The results presented in chapters 4, 5 and 6 suggest that the effect of a more female peer group has a significant effect on outcomes, but that the magnitude of this effect is generally smaller than the effect of a more able peer group.

It is interesting to compare the results obtained here for a more able peer group with the effects observed in previous literature. The previous literature examining the effect of a more able peer group on children's academic outcomes has been unable to reach a consensus on the effect of an increase in the mean ability of the peer group, with results ranging from no, or a very small significant effect (e.g. Angrist and Lang (2004) (No effect of a less able peer group introduced), Lefgren(2004b) (a 1 standard deviation increase in the peer ability measure linked with a 0.024 standard deviation increase in individuals outcomes), to a much larger effect of a magnitude of a 1 standard deviation increase in peer ability related to a 0.3 standard deviation increase in individual's achievement (e.g. Kang 2007)). Further studies have suggested effects within this range (e.g. Hoxby (2000) suggests a 1 standard deviation increase in peer ability leads to a between 0.05 and 0.14 standard deviation increase in the outcome of individual students, whilst Gibbons and Telhaj (2008) suggest that for middle achieving students, a 1 standard deviation increase in the proportion of pupils who are high achievers is related to a 0.15 standard deviation increase in their outcomes. Similarly Hanushek et al (2003) suggests a 0.1 standard deviation increase in the peer ability measure is associated with a 0.02 standard deviation increase in individuals' outcomes.

The results presented here, in contrast with some of the previous literature (Lefgren (2004b), for example) suggest that there is significant evidence of a non-trivial effect of a more able peer group on children`s outcomes, and whilst the linear-in-means specification suggests small effects, the effects observed when considering the distance a child's ability is from the ability of the peer group gives evidence of a larger effect still, with in mathematics, a large effect, of similar magnitude to that observed in Kang (2007).

Figure 12 **Distribution of month of birth of pupils**

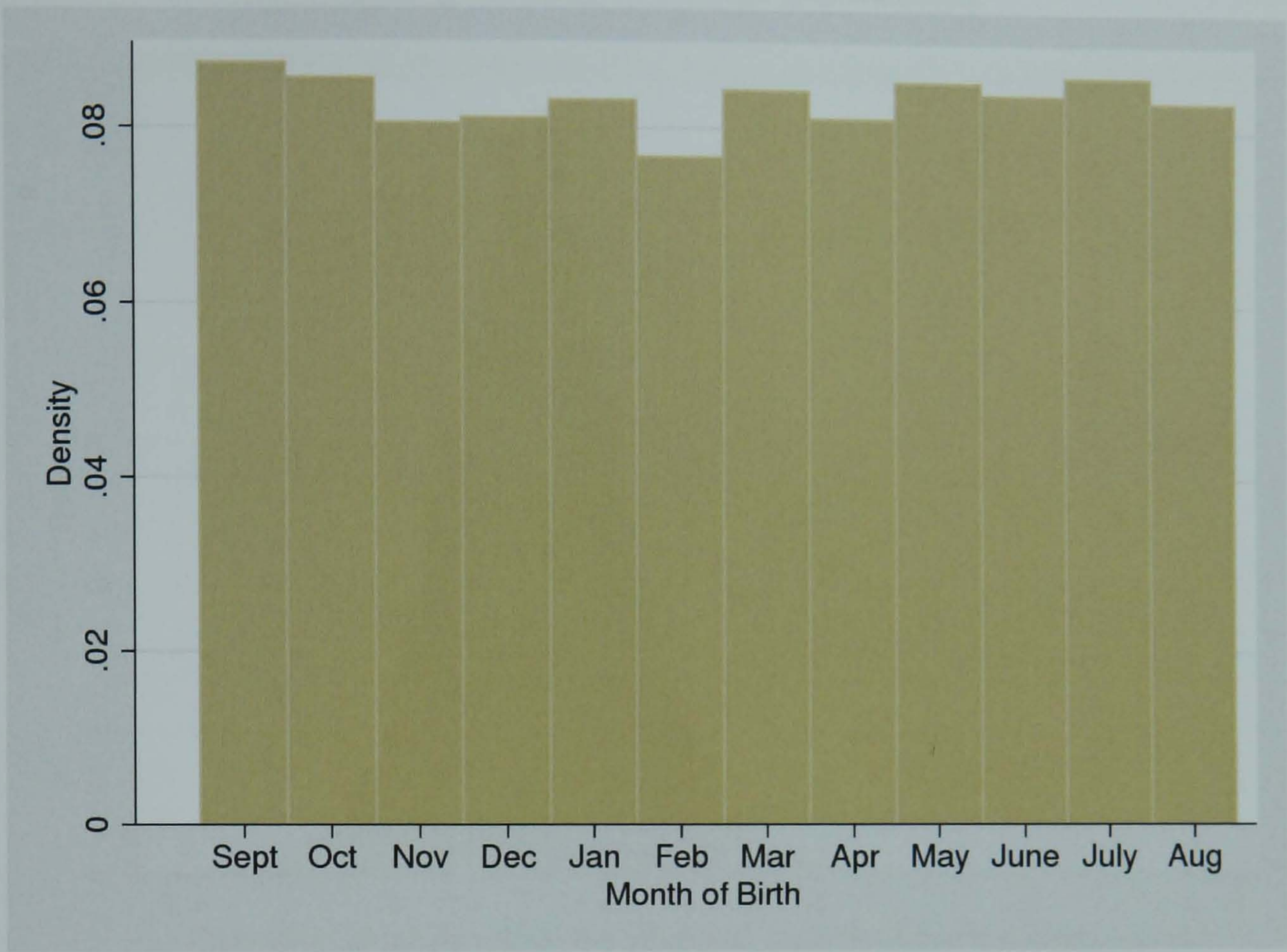


Figure 13 **Distribution of age of pupil by thirds**

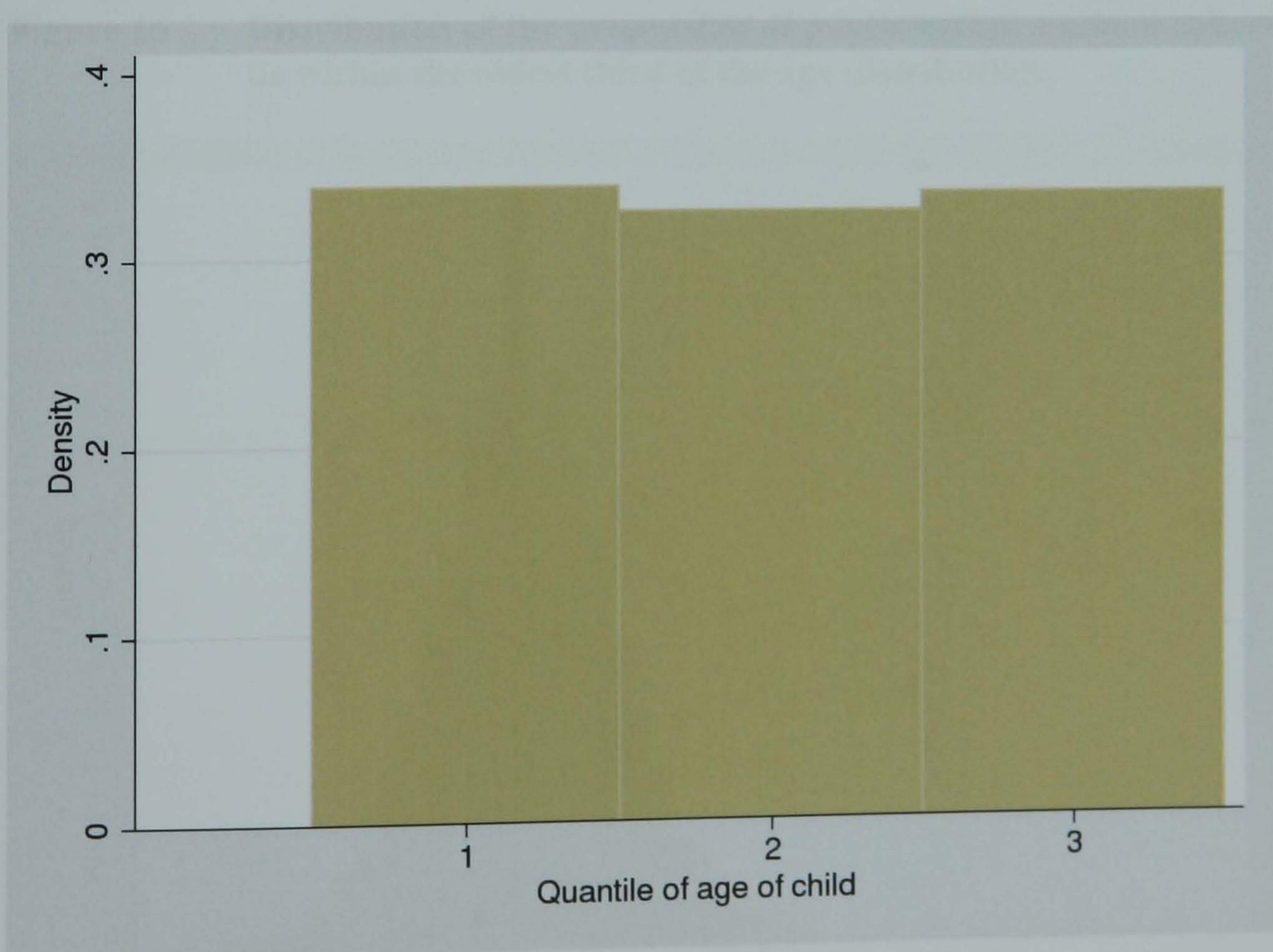
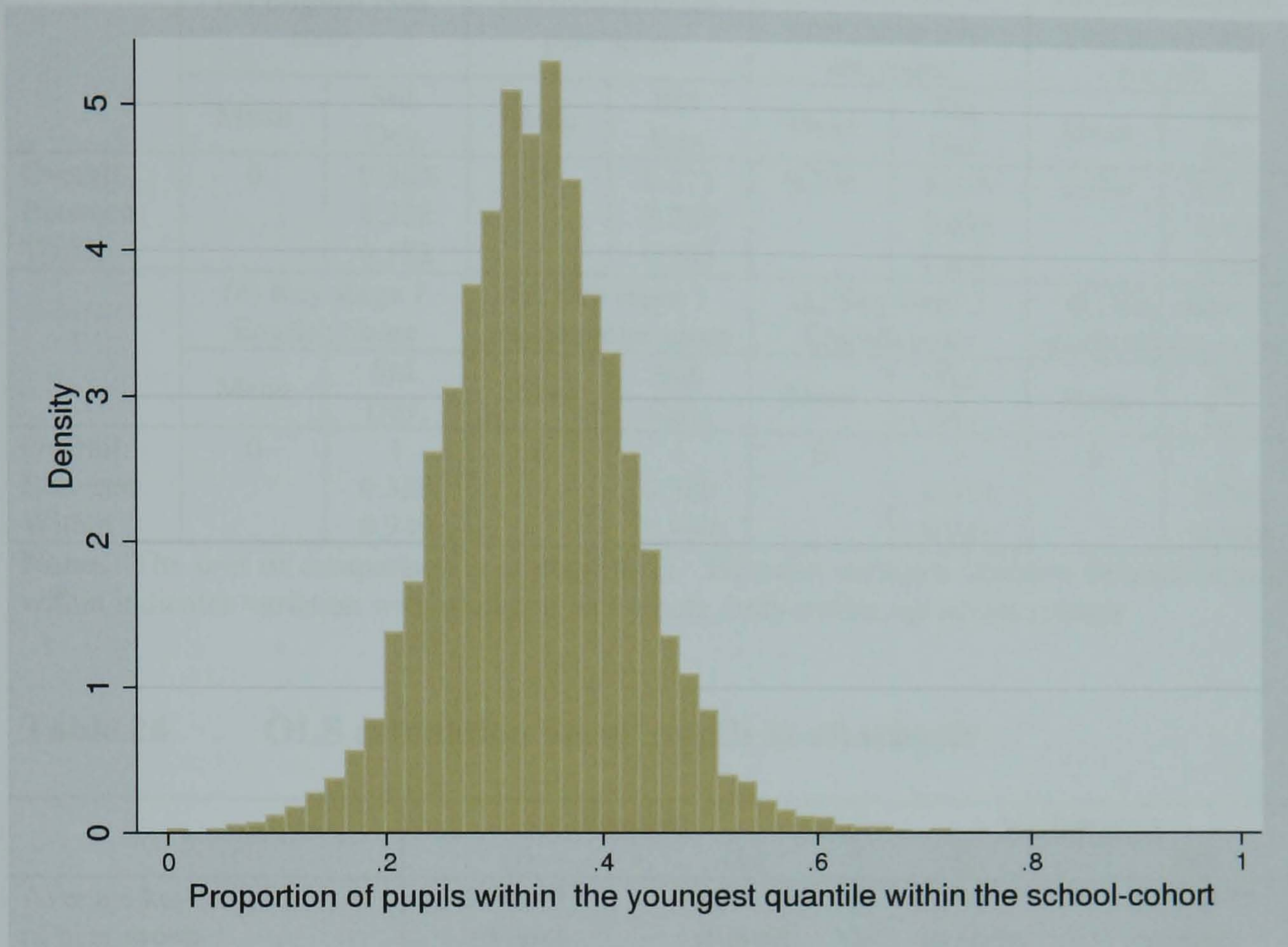
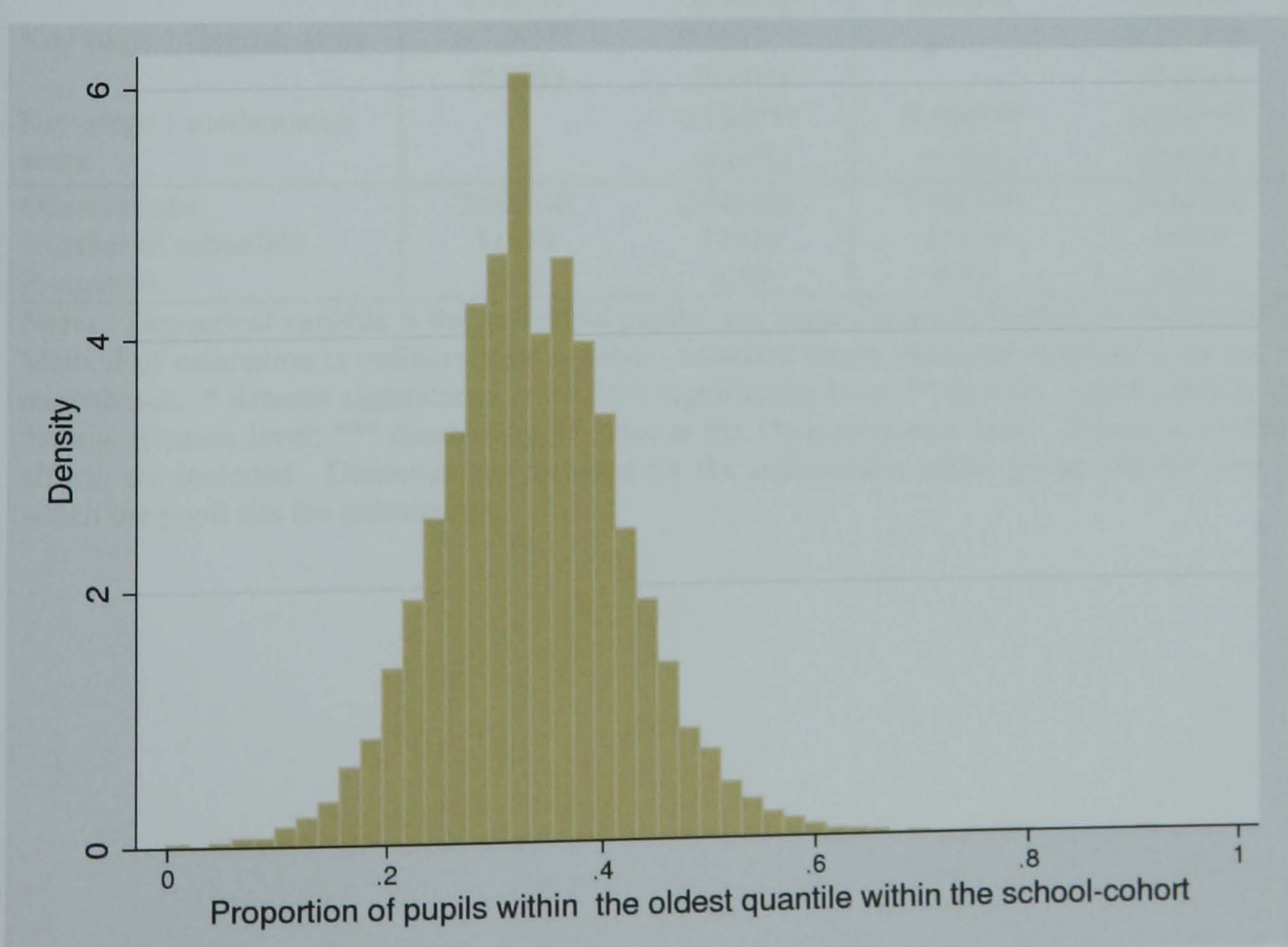


Figure 14 **Distribution of the proportion of pupils within school who are in the youngest third of the age distribution**



Notes: Unit of observation is the proportion of students within the total school cohort who is born in the youngest of three quantiles. One observation per school

Figure 15 **Distribution of the proportion of pupils within a school cohort lie within the oldest third of the age distribution**



Notes: Unit of observation is the proportion of students within the total school cohort who is born in the oldest of three quantiles. One observation per school

Table 25 Summary statistics

	(a) English Peer score		(b) Mathematics peer score		(c) Proportion of peer group who are young		(d) Proportion of peer group who are old.	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Overall	0	0.365	0	0.371	0.339	0.076	0.336	0.076
Between		0.328		0.289		0.038		0.039
Within		0.182		0.245		0.068		0.068
	(e) Key stage 1 English Score		(f) Key stage 1 mathematics score		(g) Key stage 2 English score		(h) Key stage 2 mathematics score	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Overall	0	1	0	1	0	1	0	1
Between		0.328		0.289		0.354		0.341
Within		0.949		0.960		0.941		0.945

Notes. The unit of comparison is at pupil level. Between indicates variation between schools, within indicates variation within schools as a whole, both within and across cohorts.

Table 26 OLS estimation for all pupils in all schools

	English		Mathematics	
	(i)	(ii)	(i)	(ii)
Average key stage 1 score	-0.269***	-0.280***	-0.364***	-0.329***
of peer group	(0.006)	(0.006)	(0.005)	(0.004)
Child takes free school	-0.122***	-0.114***	-0.151***	-0.094***
meals	(0.001)	(0.001)	(0.001)	(0.001)
Age of child	-0.056***	-0.103***	-0.137***	-0.180***
	(0.002)	(0.002)	(0.002)	(0.002)
Male pupil	-0.108***	-0.153***	0.098***	0.188***
	(0.001)	(0.001)	(0.001)	(0.001)
Key stage 1 English score	0.728***	0.602***		0.277***
	(0.001)	(0.001)		(0.001)
Key stage 1 mathematics		0.182***	0.720***	0.534***
score		(0.001)	(0.001)	(0.001)
Observations	2446348	2446348	2446348	2446348
Number of schoolid2	11919	11919	11919	11919
R-squared	0.56	0.58	0.53	0.57

Notes. Dependent variable is the individual pupils' key stage 2 score in English or mathematics. Method of estimation is ordinary least squares. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination.

Table 27 First stage regressions for all pupils in all schools

	English		Mathematics	
	(i)	(ii)	(i)	(ii)
Proportion of pupils in the youngest quantile	-0.200***	-0.200***	-0.237***	-0.237***
	(0.013)	(0.013)	(0.015)	(0.015)
Proportion of pupils in the oldest quantile.	0.158***	0.158***	0.188***	0.187***
	(0.013)	(0.013)	(0.015)	(0.015)
Child takes free school meals	0.006***	0.006***	0.008***	0.006***
	(0.000)	(0.000)	(0.000)	(0.000)
Age of child	-0.009***	-0.010***	-0.016***	-0.015***
	(0.000)	(0.000)	(0.001)	(0.001)
Male pupil	0.005***	0.004***	-0.001***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)
Key stage 1 English score	0.009***	0.007***		-0.011***
	(0.000)	(0.000)		(0.000)
Key stage 1 mathematics score		0.004***	0.021***	0.029***
		(0.000)	(0.000)	(0.000)
Observations	2446348	2446348	2446348	2446348
Number of schoolid2	11919	11919	11919	11919
R-squared	0.13	0.13	0.36	0.36

Notes. Dependent variable is the average of the peer group's results at key stage 1 in English or mathematics. The proportion of pupils in the youngest third and the proportion of pupils in the oldest third are introduced as excluded instruments for the average key stage 1 score of the peer group. Method of estimation is ordinary least squares. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination.

Table 28 IV estimation in all schools

	English		Mathematics	
	(i)	(ii)	(i)	(ii)
Average key stage 1 score of peer group	0.116**	0.116**	0.139***	0.162***
	(0.047)	(0.047)	(0.037)	(0.036)
Child takes free school meals	-0.125***	-0.117***	-0.155***	-0.097***
	(0.001)	(0.001)	(0.001)	(0.001)
Age of child	-0.052***	-0.098***	-0.127***	-0.171***
	(0.002)	(0.002)	(0.002)	(0.002)
Male pupil	-0.110***	-0.154***	0.099***	0.190***
	(0.001)	(0.001)	(0.001)	(0.001)
Key stage 1 English score	0.724***	0.599***		0.283***
	(0.001)	(0.001)		(0.001)
Key stage 1 mathematics score		0.181***	0.710***	0.520***
		(0.001)	(0.001)	(0.001)
Underidentification test	719.065	719.430	751.626	752.298
P-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Weak instrument test statistic	393.716	393.934	413.768	414.147
Stock Yogo Critical value	*19.93*	19.93	*19.93*	*19.93*
Hansen J statistic of overidentifying restrictions	0.616	0.536	0.263	0.072
P value	*0.4324*	*0.4642*	*0.6082*	*0.7887*
Observations	2446348	2446348	2446348	2446348
Number of schoolid2	11919	11919	11919	11919

Notes. Dependent variable is the individual pupils' key stage 2 score in English or mathematics. Method of estimation is two stage least squares. Excluded instruments for the average key stage 1 score of the peer group are the proportion of pupils who are in the youngest third and the proportion of pupils within the oldest third of the age distribution within the peer group. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination. The underidentification test is the Kleibergen Paap LM Test. The weak instrument test statistic is the Kleibergen Paap Wald F statistic. The Stock Yogo critical value is the 10% maximal IV size.

Table 29 IV estimation within small schools

	English		Mathematics	
	(i)	(ii)	(i)	(ii)
Average key stage 1 score of peer group	0.103	0.098	0.092	0.121**
	(0.074)	(0.074)	(0.061)	(0.059)
Child takes free school meals	-0.121***	-0.113***	-0.147***	-0.093***
	(0.004)	(0.004)	(0.004)	(0.004)
Age of child	-0.053***	-0.100***	-0.125***	-0.167***
	(0.004)	(0.004)	(0.005)	(0.005)
Male pupil	-0.109***	-0.153***	0.097***	0.187***
	(0.002)	(0.002)	(0.003)	(0.003)
Key stage 1 English score	0.705***	0.583***		0.270***
	(0.002)	(0.002)		(0.002)
Key stage 1 mathematics score		0.178***	0.683***	0.501***
		(0.002)	(0.003)	(0.003)
Underidentification test statistic	218.222	218.224	259.362	259.412
P-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Weak instrument test statistic	121.640	121.639	148.077	148.089
Stock Yogo Critical value	*19.93*	*19.93*	*19.93*	*19.93*
Hansen J statistic of overidentifying restrictions	0.049	0.074	0.042	0.095
P value	*0.8254*	*0.7852*	*0.8369*	*0.7584*
Observations	326455	326455	326455	326455
Number of schoolid2	3056	3056	3056	3056

Notes. Dependent variable is the individual pupils' key stage 2 score in English or mathematics. Method of estimation is two stage least squares. Excluded instruments for the average key stage 1 score of the peer group are the proportion of pupils who are in the youngest third and the proportion of pupils within the oldest third of the age distribution within the peer group. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination. A small school is defined as a school that has 30 or fewer pupils in every observed cohort. The underidentification test is the Kleibergen Paap LM Test. The weak instrument test statistic is the Kleibergen Paap Wald F statistic. The Stock Yogo critical value is the 10% maximal IV size.

Table 30 IV estimation for English considering the individual pupil's difference in ability compared with the ability of the peer group

	(a) Lowest quartile of distance from average peer key stage 1 score		(b) Second quartile of distance from average peer key stage 1 score		(c) Third quartile of distance from average peer key stage 1 score		(d) Highest quartile of distance from average peer key stage 1 score	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
Average English key stage 1 score of peer group	0.324***	0.315***	0.424***	0.455***	0.534***	0.541***	0.201***	0.186***
	(0.068)	(0.068)	(0.124)	(0.124)	(0.112)	(0.112)	(0.063)	(0.063)
Child takes free school meals	-0.107***	-0.102***	-0.129***	-0.120***	-0.124***	-0.116***	-0.130***	-0.122***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Age of child	-0.095***	-0.145***	-0.088***	-0.139***	-0.065***	-0.102***	-0.000	-0.033***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Male pupil	-0.084***	-0.130***	-0.101***	-0.147***	-0.112***	-0.153***	-0.126***	-0.160***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Key stage 1 English score	0.465***	0.351***	0.583***	0.416***	0.447***	0.324***	0.558***	0.473***
	(0.006)	(0.005)	(0.061)	(0.061)	(0.050)	(0.050)	(0.012)	(0.011)
Key stage 1 mathematics score		0.187***		0.189***		0.152***		0.145***
		(0.001)		(0.001)		(0.001)		(0.001)
Underidentification test statistic	665.514	666.163	514.607	514.951	513.545	513.982	628.759	628.831
P-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Weak instrument test statistic	362.327	362.693	271.784	271.984	269.111	269.350	340.792	340.819
Stock Yogo Critical value	*19.93*	*19.93*	*19.93*	*19.93*	19.93	*19.93*	*19.93*	19.93
Hansen J statistic of overidentifying restrictions	0.024	0.070	0.152	0.135	0.928	0.538	0.454	0.425
P value	*0.8759*	*0.7913*	*0.6967*	*0.7138*	*0.3353*	*0.4632*	*0.5003*	*0.5142*
Observations	611755	611755	611431	611431	611581	611581	611581	611581
Number of schoolid2	11919	11919	11919	11919	11919	11919	11919	11919
R-squared								

Notes. Dependent variable is the individual pupils' key stage 2 score in English. Method of estimation is two stage least squares. Excluded instruments for the average key stage 1 score of the peer group are the proportion of pupils who are in the youngest third and the proportion of pupils within the oldest third of the age distribution within the peer group. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination. Specifications (a)-(d) are defined by the distribution of the distance of individual pupils' key stage 1 score from the average key stage 1 score of their peer group, so (a) is the furthest below the peer ability whilst (d) is the furthest above the peer ability score. The underidentification test is the Kleibergen Paap LM Test. The weak instrument test statistic is the Kleibergen Paap Wald F statistic. The Stock Yogo critical value is the 10% maximal IV size.

Table 31 IV estimation for English considering the individual pupil's difference in ability compared with the ability of the peer group in small schools

	(a) Lowest quartile of distance from average peer key stage 1 score		(b) Second quartile of distance from average peer key stage 1 score		(c) Third quartile of distance from average peer key stage 1 score		(d) Highest quartile of distance from average peer key stage 1 score	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
Average English key stage 1 score of peer group	0.387***	0.363***	0.589*	0.661**	0.520**	0.522**	0.169	0.149
	(0.117)	(0.116)	(0.306)	(0.305)	(0.232)	(0.231)	(0.107)	(0.107)
Child takes free school meals	-0.113***	-0.105***	-0.120***	-0.112***	-0.123***	-0.115***	-0.122***	-0.115***
	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Age of child	-0.096***	-0.146***	-0.097***	-0.145***	-0.076***	-0.116***	0.007	-0.026***
	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)
Male pupil	-0.095***	-0.140***	-0.094***	-0.139***	-0.110***	-0.152***	-0.120***	-0.155***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Key stage 1 English score	0.433***	0.326***	0.378**	0.185	0.380***	0.256*	0.551***	0.464***
	(0.014)	(0.013)	(0.191)	(0.190)	(0.132)	(0.131)	(0.028)	(0.028)
Key stage 1 mathematics score		0.179***		0.183***		0.158***		0.147***
		(0.004)		(0.004)		(0.003)		(0.003)
Underidentification test statistic	198.471	198.400	112.064	112.137	140.211	140.269	189.339	189.042
P-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Weak instrument test statistic	110.318	110.257	58.630	58.674	74.003	74.033	104.093	103.908
Stock Yogo Critical value	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*
Hansen J statistic of overidentifying restrictions	0.101	0.155	0.702	0.502	0.158	0.028	0.018	0.005
P value	*0.7503*	*0.6934*	*0.4021*	*0.4786*	*0.6907*	*0.8676*	*0.8926*	*0.9442*
Observations	83497	83497	77724	77724	83145	83145	82089	82089
Number of schoolid2	3056	3056	3056	3056	3056	3056	3056	3056

Notes. Dependent variable is the individual pupils' key stage 2 score in English. Method of estimation is two stage least squares. Excluded instruments for the average key stage 1 score of the peer group are the proportion of pupils who are in the youngest third and the proportion of pupils within the oldest third of the age distribution within the peer group. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination. Specifications (a)-(d) are defined by the distribution of the distance of individual pupils' key stage 1 score from the average key stage 1 score of their peer group, so (a) is the furthest below the peer ability whilst (d) is the furthest above the peer ability score. A small school is defined as a school that has 30 or fewer pupils in every observed cohort. The underidentification test is the Kleibergen Paap LM Test. The weak instrument test statistic is the Kleibergen Paap Wald F statistic. The Stock Yogo critical value is the 10% maximal IV size.

Table 32 IV estimation for mathematics considering the individual pupil's difference in ability compared with the ability of the peer group

	(a) Lowest quartile of distance from average peer key stage 1 score		(b) Second quartile of distance from average peer key stage 1 score		(c) Third quartile of distance from average peer key stage 1 score		(d) Highest quartile of distance from average peer key stage 1 score	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
Average mathematics key stage 1 score of peer group	0.488***	0.449***	1.233***	1.217***	0.507***	0.550***	0.178***	0.200***
	(0.065)	(0.062)	(0.172)	(0.167)	(0.101)	(0.099)	(0.045)	(0.044)
Child takes free school meals	-0.120***	-0.073***	-0.152***	-0.098***	-0.159***	-0.104***	-0.174***	-0.112***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Age of child	-0.193***	-0.221***	-0.194***	-0.231***	-0.143***	-0.186***	-0.034***	-0.094***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
Male pupil	0.121***	0.192***	0.096***	0.191***	0.089***	0.187***	0.100***	0.184***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)
Key stage 1 English score	0.462***	0.316***	-0.067	-0.263**	0.340***	0.142***	0.303***	0.171***
	(0.007)	(0.007)	(0.113)	(0.110)	(0.051)	(0.050)	(0.014)	(0.014)
Key stage 1 mathematics score		0.231***		0.285***		0.284***		0.291***
		(0.002)		(0.002)		(0.002)		(0.001)
Underidentification test statistic	649.480	650.275	254.772	255.368	367.435	366.985	550.475	550.307
P-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Weak instrument test statistic	353.246	353.725	132.343	132.667	192.276	192.030	295.237	295.132
Stock Yogo Critical value	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*
Hansen J statistic of overidentifying restrictions	0.056	0.046	1.092	0.688	0.645	0.242	0.170	0.035
P value	*0.8135*	*0.8294*	*0.2961*	*0.4068*	*0.4219*	*0.6228*	*0.6804*	*0.8517*
Observations	612426	612426	610949	610949	611633	611633	611337	611337
Number of schoolid2	11919	11919	11913	11913	11879	11879	11909	11909

Notes. Dependent variable is the individual pupils' key stage 2 score in mathematics. Method of estimation is two stage least squares. Excluded instruments for the average key stage 1 score of the peer group are the proportion of pupils who are in the youngest third and the proportion of pupils within the oldest third of the age distribution within the peer group. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination. Specifications (a)-(d) are defined by the distribution of the distance of individual pupils' key stage 1 score from the average key stage 1 score of their peer group, so (a) is the furthest below the peer ability whilst (d) is the furthest above the peer ability score. The underidentification test is the Kleibergen Paap LM Test. The weak instrument test statistic is the Kleibergen Paap Wald F statistic. The Stock Yogo critical value is the 10% maximal IV size.

Table 33 **IV estimation for mathematics considering the individual pupil's difference in ability compared with the ability of the peer group in small schools**

	(a) Lowest quartile of distance from average peer key stage 1 score		(b) Second quartile of distance from average peer key stage 1 score		(c) Third quartile of distance from average peer key stage 1 score		(d) Highest quartile of distance from average peer key stage 1 score	
	(i)	(ii)	(i)	(ii)	(i)	(ii)	(i)	(ii)
Average mathematics key stage 1 score of peer group	0.506***	0.481***	1.523***	1.481***	0.629***	0.654***	0.133	0.167**
	(0.121)	(0.117)	(0.436)	(0.422)	(0.235)	(0.228)	(0.084)	(0.083)
Child takes free school meals	-0.121***	-0.078***	-0.146***	-0.098***	-0.142***	-0.088***	-0.163***	-0.107***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Age of child	-0.180***	-0.210***	-0.188***	-0.225***	-0.142***	-0.183***	-0.045***	-0.099***
	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.008)	(0.007)	(0.006)
Male pupil	0.120***	0.189***	0.089***	0.184***	0.086***	0.183***	0.100***	0.182***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
Key stage 1 English score		0.219***		0.271***		0.278***		0.277***
		(0.004)		(0.005)		(0.004)		(0.004)
Key stage 1 mathematics score	0.411***	0.274***	-0.447	-0.609*	0.193	0.012	0.311***	0.176***
	(0.020)	(0.020)	(0.331)	(0.322)	(0.142)	(0.138)	(0.032)	(0.032)
Underidentification test statistic	199.106	199.432	52.120	52.287	95.817	95.761	170.715	170.552
P-value	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*	*0.0000*
Weak instrument test statistic	110.759	110.948	26.957	27.048	50.079	50.052	93.503	93.405
Stock Yogo Critical value	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*	*19.93*
Hansen J statistic of overidentifying restrictions	0.093	0.285	0.551	0.367	0.235	0.136	1.267	0.837
P value	*0.7604*	*0.5936*	*0.4579*	*0.5444*	*0.6276*	*0.7128*	*0.2604*	*0.3601*
Observations	81877	81877	80131	80131	85469	85469	78977	78977
Number of schoolid2	3056	3056	3054	3054	3049	3049	3054	3054

Notes. Dependent variable is the individual pupils' key stage 2 score in mathematics. Method of estimation is two stage least squares. Excluded instruments for the average key stage 1 score of the peer group are the proportion of pupils who are in the youngest third and the proportion of pupils within the oldest third of the age distribution within the peer group. Standard errors, clustered at school level are in parentheses. * denotes significance at the 10% significance level; ** denotes significance at the 5% significance level; *** denotes significance at the 1% significance level. School level fixed effects are included. Dummies are included for the individual's ethnic group and the year in which the pupil sits the examination. Specifications (a)-(d) are defined by the distribution of the distance of individual pupils' key stage 1 score from the average key stage 1 score of their peer group, so (a) is the furthest below the peer ability whilst (d) is the furthest above the peer ability score. A small school is defined as a school that has 30 or fewer pupils in every observed cohort. The underidentification test is the Kleibergen Paap LM Test. The weak instrument test statistic is the Kleibergen Paap Wald F statistic. The Stock Yogo critical value is the 10% maximal IV size.

Chapter 7 Conclusions

The introduction of a quasi-market into schools in the 1980s and 1990s in England allowed much more transparency to enable parents to choose which school to send their children to, and to allow more accountability within education. A key innovation in the quasi-market was the introduction of league tables in 1992, initially offering only summary of the performance of the students, but have evolved since then to include measures of value added within the school. If these market systems are improve the effectiveness of schools, it is important to identify the factors which help to improve pupils' outcomes.

The Coleman Report (Coleman et al (1966)) identified factors that affected a child's outcomes in schools, and suggested that children's outcomes are influenced, in decreasing magnitudes, by their own demographics, the demographics of their peers and finally by the facilities, teachers and curriculum of the school. Whilst Kramarz et al (2009) suggest that a child's peers have less of an effect than schools, it is still important that we consider both the effects of a child's peer groups and the effectiveness of schools.

Whilst Coleman's analysis raised the profile of the effect of a child's peers on their outcomes, it does have methodological flaws. Because parents sort their children into schools with pupils of similar ability (see, for example Burgess et al (2007), Urquiola (2005)), there are difficulties in estimating the effect of the peer group on individuals' outcomes. If the ability, say, of the peer group affects individuals' outcomes, then this sorting of pupils into schools with peers who 'look like' them in terms of ability will necessarily make this peer ability measure endogenous (see, for example Manski (1993)).

In this thesis, I have considered both the effect of the gender make-up of a child's peer groups on outcomes, and the effect of a more able peer group on pupils' outcomes at ages 11 and 16, and have shown that both of these properties of the peer group have a significant and non-trivial effect on students' outcomes. Chapter 4 examines the gender make-up of the peer group, using exogenous variation in the proportion of the school cohort which is female to identify the effect. The results I have found for mathematics and science are in line with those shown by Hoxby (2000) and Lavy and Schlosser (2007), with a significant positive effect of a more female peer group. However, my results also suggest that a more female peer group has a negative effect on boys' outcomes in English, with an effect that is persistent at all levels of assessment, except key stage 2. Furthermore, girls are *not* negatively impacted by a more female peer group. The results presented in Table 7 in chapter 4 suggest that these effects are largely linear. In terms of policy, these results tend to suggest that boys might benefit from being educated in single sex classes. However, as noted previously, since the effects are identified using year on year changes in the proportion of the cohort that is female and given that there are no single sex cohorts included, generalising to single sex classrooms is difficult. Conversely, all pupils would benefit from being taught mathematics and science with a larger proportion of girls. However, here gains for boys would also be countered with losses for girls, making policy implications difficult to infer. A natural extension to this chapter would be to consider children who move from mixed education to single sex education. However, as with all studies that compare education in single sex schools with those in mixed schools, selection issues would make the identification issues difficult as there are numerous unobserved heterogeneities between single sex and mixed schools, such as school ethos and parental beliefs.

Chapter 5 offered a novel technique for estimating the effect of a more able peer groups at age 16, taking advantage of a dataset containing full data for which classes pupils were taught in for English, mathematics and science for a small subset of schools. The method of estimation considered schools that used credibly random allocation of pupils to classrooms within GCSE tiers, and estimated significant positive effects of a more able peer group at age 16. As a validation technique, IV methods following a methodology developed by Lefgren (2004b) gained similar estimates of the magnitude of the effect of a more able peer group to those estimated from the credibly random sample. These effects suggest that within tiers, pupils benefit from an increase in the ability of their peer group. Further, the results suggest that the highest ability pupils in the lower tier would perform better if they were placed in the higher tier classroom, but the introduction of lower ability pupils into the higher ability classrooms would have the effect of lowering the mean ability of the peer group, and consequently the achievement of the other pupils in the class. Since this classroom level dataset only contains a small subset of schools, it is difficult to conclude whether this is a zero sum game, or whether an increase in the ability of the peer group would have the same effect on low ability students within tier as with high ability students.

Chapter 6 offers a further technique for estimating the effect of a more able peer group at age 11, taking advantage of the proportion of pupils who are in the oldest third of the cohort and the proportion in the youngest third as an instrument for the ability of the peer group. The results obtained suggest a similar magnitude of effect from a more able peer group as those obtained in chapter 5, again suggesting that a more able peer group has a significant and non-trivial positive effect. Further, this chapter suggests that the size of the difference between a child's own prior ability and the ability of the peer group is important for the size of the effect, with children whose ability is close to the average of the peer group experiencing a larger effect from the peer influence than those whose ability is far away from the ability of the peer group. The strategy employed in this analysis has the obvious drawback that due to the constraints of the data-set, it is difficult to correctly identify the exact classroom peer group, and the identification of effects relies on the make up of the entire within-school cohort. The results from chapters 5 and 6 offer a new estimate of the size of the effect due to the ability of the peer group, which are larger than several previous estimates, suggesting a 1 standard deviation increase in the mean ability of the peer group being associated with a between 0.05 and 0.4 standard deviation increase in individual's outcomes. The effects from a more able peer group are also larger than those estimated in chapter 4 for the effect of a more female peer group, suggesting that the ability of the peer group is more important for influencing pupils' outcomes than the gender mix of the school.

In order to assess the importance of these peer effects in the educational production function, it is possible to compare the estimated size of the effect of an increase in the prior achievement of the peer group with the effect of other characteristics within this specification[52]. Children who are eligible for free school meals perform approximately 0.6 standard deviations worse on average than non free school meals students. Being one month older within the year improves outcomes by between 0.01 and 0.05 standard deviations, but this advantage decreases from key stage 1 to key stage 4. On average, the effect is approximately a 0.03 standard deviation increase for children who are a month older. Considering ethnicity, there is a range of 0.6 standard deviations between the results of students of Chinese origin, the highest achieving of the major ethnic groups, and the results of students of Black Caribbean origin, the lowest achieving of the major ethnic groups on average. In English, boys perform approximately 0.3 standard deviations lower than girls, whilst in maths and science, there is little difference between girls and boys.

The results in chapter 4 for English suggest that a 10 percentage point increase in the proportion of pupils who are female is associated with a between 0.01 and 0.03 standard deviation decrease in boys' scores, whilst boys and girls perform 0.01 standard deviations better in maths and science from a 10 percentage point increase in the proportion of pupils who are female. These effects are small, but are of a similar magnitude to the pupils being up to one month older within the school year.

The results for the ability of the peer group are of a larger magnitude than those for a change in the gender make-up of the peer group. From chapter 5, in English, a 1 standard deviation increase in the prior attainment of the peer group is associated with a between 0.07 and 0.11 standard deviation increase in pupils outcomes at age 16. This is approximately one third of the difference between boys and girls at English, or alternatively equivalent to the pupils being between 3 and 4 months older within the school year. For mathematics, a larger effect again was seen with a 1 standard deviation increase in the peer group's prior attainment associated with a 0.2 standard deviation increase in the individuals' outcomes, or approximately one third of the difference between

[52] These figures are taken from OLS regressions of key stage performance on, FSM status, year of exam, age of student, gender, ethnic group, whether they have English as a first language, and a dummy for the level of the key stage. These coefficients are unconditional on prior attainment, as this is strongly correlated with demographics I am interested in.

FSM and non-FSM students, or equivalent to the pupils being approximately 7 months older within the school year. Considering the regressions broken down by difference in ability between individuals and their peer group, an effect of a larger magnitude is seen. In English, for those close to the ability of their peer group, the effect of a 1 standard deviation increase in the prior attainment of the peer group is approximately half of the magnitude of the difference between girls and boys outcomes in English, or equivalently the pupils being between 5 and 6 months older. The estimates for children with a large difference between their attainment and the attainment of their peer group are of a smaller magnitude, equivalent to being between 3 and 4 months older. For mathematics, a 1 standard deviation increase in the peer ability for pupils who have attainment just below the attainment of their peer group has an effect equivalent to approximately half of the difference between the best and worst ethnic groups, or alternatively approximately the pupils being between 10 and 12 months older. The results for other pupils in mathematics suggest a smaller effect of a 1 standard deviation increase in peer ability, equivalent to the pupils being approximately 4 months older within the school year.

Overall, this thesis suggests that the overall magnitude of the effect of a child's peer group is economically significant, albeit small. These effects are dwarfed by the differences in outcomes associated with ethnicity, but are of a similar magnitude to being between 1 and 6 months older within the school-year.

A future extension to the work on peer effects might be to consider at what level the peer effects are occurring, whether at the school level, classroom level or within small-group level. In chapter 5, I have explicitly estimated the effects of a more able peer group within classrooms in secondary schools, and have obtained similar results to those estimated at school level in primary schools. It would be interesting to observe what effect peer characteristics within a working group or a friendship group has on outcomes. Mora and Oreopoulos (2009) offer an interesting model that could be built upon to consider the effect of peer influence within friendship groups. They consider which pupils within classes are considered friends based on questionnaire results, and their results suggest that when it comes to intention whether to drop out of school, these effects occurred only when both pupils named each other as friends. For example, if this methodology were applied to the question in chapter 4, it would be interesting to see the extent the effect that more close friends who are female has on both boys and girls outcomes in schools, and it is conceivable this effect could be different from the effect observed at the school or classroom level.

Education remains a key priority for governments, as it offers the potential to lower social inequality within the economy and to benefit the nation as a whole by raising the skills of the population. However, analysing education and making informed policy decisions has been difficult due to the complex nature of schooling and the educational production function. In recent years, these difficulties have been eased by the availability of new sources of data covering pupils' outcomes and their characteristics. This thesis has exploited the availability of these new data sources to contribute new evidence to the field of the economics of education.

References

AMMERMUELLER, A., and PISCHKE, J. S. (2009): "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study," *Journal of Labor Economics*, 27(3), 315-348.

ANGRIST, J. D., and KRUEGER, A. B. (1991): "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106(4), 979-1014.

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., GREGG, P., PROPPER, C., SLATER, H., and WILSON, D. (2004) "Evaluating the performance-related pay for teachers in England," CMPO Working Paper 04/113 Centre for Market and Public Organisation, University of Bristol

— (2009): "Evaluating the impact of performance-related pay for teachers in England," *Labour Economics*, 16(3), 251-261.

ATKINSON, A., GREGG, P., and MCCONNELL, B. (2006) "The Result of 11 Plus Selection: An Investigation into Opportunities and Outcomes for Pupils in Selective LEAs," CMPO Working Paper 06/150, Centre for Market and Public Organisation, University of Bristol

BAUM, C. F., SCHAFFER, M. E., and STILLMAN, S. (2007): "ivreg2: Stata module for extended instrumental variables/2SLS, GMM and AC/HAC, LIML, and k-class regression," *Boston College Department of Economics, Statistical Software Components S425401. Downloadable from http://ideas.repec.org/c/boc/bocode/s425401.html.*

BBC (2008): "Summer-born to start school later," http://news.bbc.co.uk/1/hi/education/7178969.stm accessed on 06/07/2009

— (2009): "Is this the worst day of the year to be born?," http://news.bbc.co.uk/1/hi/magazine/8227268.stm accessed on 11/09/09

BELL, J. F., and DANIELS, S. (1990): "Are Summer-Born Children Disadvantaged? The Birthdate Effect in Education," *Oxford Review of Education*, 16(1), 67-80.

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," mimeo Public Policy Institute of California, University of California

BORG, M. G., and FALZON, J. M. (1995): "Birth Date and Sex Effects on the Scholastic Attainment of Primary Schoolchildren: A Cross-Sectional Study," *British Educational Research Journal*, 21(1), 61-74.

BOUND, J., JAEGER, D. A., and BAKER, R. M. (1995): "Problems with Instrumental Variables Estimation When the Correlation between the Instruments and the Endogenous Explanatory Variable Is Weak," *Journal of the American Statistical Association*, 90(430), 443-450.

BRADLEY, S., and TAYLOR, J. (2007) "Do Peers Matter? Estimation of Peer Effects from Pupil Mobility Between Schools," LERG discussion paper 13, University of Lancaster

BURGESS, S., MCCONNELL, B., PROPPER, C., and WILSON, D. (2004): "Girls rock, boys roll: An analysis of the age 14-16 gender gap in English schools," *Scottish Journal of Political Economy*, 51(2), 209-229.

— (2007): "The Impact of School Choice on Sorting by Ability and Socioeconomics Factors in English Secondary Education" in *Schools and the Equal Opportunity Problem*, edited by L. Woessmann, and P. E. Peterson. Cambridge, Massachusetts: The MIT Press, 273-291.

BURKE, M., and SASS, T. (2008) "Classroom Peer Effects and Student Achievement," Working Paper 18, National Center for Analysis of Longitudinal Data in Education Research

CAMPBELL, P. B., and SANDERS, J. (2002): "Challenging the system: Assumptions and data behind the push for single-sex schooling" in *Doing Gender in Policy and Practice: Perspectives on Single Sex and Coeducational Schooling*, edited by A. Datnow, and L. Hubbard. New York: Routledge Falmer, 31-46.

CARRELL, S. E., FULLERTON, R. L., and WEST, J. E. (2009): "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27(3), 439-464.

CHEVALIER, A., DOLTON, P., and LEVACIC, R. (2005): "School and Teacher Effectiveness" in *What's the Good of education*, edited by S. Machin, and A. Vignoles. Princeton: Princeton University Press.

CLARK, C. (2005): "Do peers matter? Evidence from the sixth grade experiment," Mimeo, University of Kentucky.

COLEMAN, J. S., CAMPBELL, E. Q., J., H. C., MCPARTLAND, J., M., M. A., D, W. F., and L, Y. R. (1966) "Equality of Educational Opportunity," OE-38001 and supp., US Department of Health, Education & Welfare, Washington, DC.

CRAWFORD, C., DEARDEN, L., and MEGHIR, C. (2007) "When you are born matters: the impact of date of birth on child cognitive outcomes in England," Discussion Paper 93, LSE Centre for the Economics of Education

DEE, T. S. (2007): "Teachers and the Gender Gaps in Student Achievement," *Journal of Human Resources*, 42(3), 528-54.

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

DING, W. L., and LEHRER, S. F. (2007): "Do peers affect student achievement in China's secondary schools?," *Review of Economics and Statistics*, 89(2), 300-312.

FEINSTEIN, L., and SYMONS, J. (1999): "Attainment in secondary school," *Oxford Economic Papers-New Series*, 51(2), 300-321.

FORD, W. C. L., NORTH, K., TAYLOR, H., FARROW, A., HULL, M. G. R., GOLDING, J., and TEAM, A. S. (2000): "Increasing paternal age is associated with delayed conception in a large population of fertile couples: evidence for declining fecundity in older men," *Human Reproduction*, 15(8), 1703-1708.

FRANCIS, B. (2000): *Boys, Girls and Achievement: Addressing the Classroom Issues*. London: Routledge Falmer.

GIBBONS, S., and MACHIN, S. (2003): "Valuing English Primary Schools," *Journal of Urban Economics*, 53(2), 197-219.

GIBBONS, S., and TELHAJ, S. (2008) "Peers and achievement in England's secondary schools," SERC Discussion Paper 1, Spatial Economics Research Centre

GORARD, S., REES, G., and SALISBURY, J. (2001): "Investigating the patterns of differential attainment of boys and girls at school," *British Educational Research Journal*, 27(2), 125-139.

HALLINAN, M. T. (1990): "The effects of Ability Grouping in Secondary Schools: A Response to Slavin's Best-Evidence Synthesis," *Review of Educational Research*, 60(3), 501-504.

HALLINAN, M. T., and SORENSEN, A. B. (1985): "Ability Grouping and Student Friendships," *American Educational Research Journal*, 22(4), 485-499.

— (1987): "Ability Grouping and Sex-Differences in Mathematics Achievement," *Sociology of Education*, 60(2), 63-72.

HANSEN, Z., OWAN, H., and PAN, J. (2006) "The Impact of Group Diversity on Performance and Knowledge Spillover - An Experiment in a College Classroom," NBER Working Paper 12551, National Bureau of Economic Research

HANUSHEK, E. A., KAIN, J. F., MARKMAN, J. M., and 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.

HERBERT, D. T., and THOMAS, C. J. (1998): "School performance, league tables and social geography," *Applied Geography*, 18(3), 199-223.

HOEL, J., PARKER, J., and RIVENBURG, J. (2006) "A Test for Classmate Peer Effects in Higher Education," Mimeo Reed College

HOXBY, C. (2000) "Peer Effects in the Classroom: Learning from Gender and Race Variation," NBER Working Paper 7867, National Bureau of Economic Research

— (2002): "The Power of Peers," *Education Next*, 2(2), 57-63.

HOXBY, C., and WEINGARTH, G. (2005) "Taking Race out of the Equation: School Reassignment and the structure of peer effects," Mimeo, Harvard University

IMBERMAN, S., KUGLER, A. D., and SACERDOTE, B. (2009) "Katrina's Children: Evidence on the Structure of Peer effects from Hurricane Evacuees," NBER Working Paper 15291, National Bureau of Economic Research

JACKSON, C. (2002): "Can single-sex classes in co-educational schools enhance the learning experiences of girls and/or boys? An exploration of pupils' perceptions," *British Educational Research Journal*, 28(1), 37-48.

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. Washington: National Academic Press.

JONES, P., and ELIAS, P. (2006) "Administrative data as a research resource: a selected audit," Economic & Social Research Council Regional Review Board Report 43/06, Warwick Institute for Employment Research

KANG, C. H. (2007): "Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea," *Journal of Urban Economics*, 61(3), 458-495.

KLEIBERGEN, F., and PAAP, R. (2006): "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 133(1), 97-126.

KRAMARZ, F., MACHIN, S., and OUAZAD, A. (2009) "What Makes a Test Score? The Respective Contribution of Pupils, Schools and Peers in Achievement in English Primary Schools," CEE Discussion Paper 102, Centre for the Economics of Education

KUBITSCHEK, W. N., and HALLINAN, M. T. (1998): "Tracking and students' friendships," *Social Psychology Quarterly*, 61(1), 1-15.

LAVY, V., PASSERMAN, D., and SCHLOSSER, A. (2008) "Inside the Black Box of Ability Peer Effects:Evidence from Variation in Low Achievers in the Classroom," NBER Working Paper 14415, National Bureau of Economic Research

LAVY, V., and SCHLOSSER, A. (2007) "Mechanisms and Impacts of Gender Peer Effects at School," NBER Working Paper 13292, National Bureau of Economic Research

LEFGREN, L. (2004a) "Educational Peer Effects and the Chicago Public Schools," mimeo, Brigham Young University

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

MACHIN, S., and MCNALLY, S. (2005): "Gender and student achievement in English schools," *Oxford Review of Economic Policy*, 21(3), 357-372.

MACLEOD, D. (2005): "Drive to recruit more male primary teachers," http://education.guardian.co.uk/primaryeducation/story/0,,1591234,00.html accessed on 21/08/09

MALACOVA, E. (2007): "Effect of single-sex education on progress in GCSE," *Oxford Review of Education*, 33(2), 233-259.

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

MARSH, H. W., and ROWE, K. J. (1996): "The effects of single-sex and mixed-sex mathematics classes within a coeducational school: A reanalysis and comment," *Australian Journal of Education*, 40(2), 147-161.

MAURIN, E., MCNALLY, S., and OUAZAD, A. (2005) "Peer Effects in English Primary Schools," Economics of Education and Education Policy in Europe Conference, Uppsala University and the Institute for Labor Market Policy Evaluation (IFAU)

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. Washington, DC: Brookings Institution Press, 79-102.

MCEWAN, P. J. (2003): "Peer effects on student achievement: evidence from Chile," *Economics of Education Review*, 22(2), 131-141.

MCINTOSH, S. (2006): "Further analysis of the returns to academic and vocational qualifications," *Oxford Bulletin of Economics and Statistics*, 68(2), 225-251.

MELKONIAN, M., and IEROKIPIOTIS, E. (1997): "The effect of age-position and sex on academic performance: a study of secondary schools in Cyprus," *Educational Research*, 39(3), 355-363.

MORA, T., and OREOPOULOS, P. (2009) "Peers effects on high school aspirations: non-reciprocating relations," Mimeo, Universitat Internacional de Catalunya, available from http://www.ugr.es/~montero/XVIeep/50.pdf

PARKER, J., GRANT, J., CROUTER, J., and RIVENBURG, J. (2008) "Classmate Peer Effects: Evidence from Core Courses at Three Colleges," Mellon Workshop on Evaluating Teaching and Learning at Liberal Arts Colleges, Wellesley College

RIVKIN, S. G., HANUSHEK, E. A., and KAIN, J. F. (2005): "Teachers, schools, and academic achievement," *Econometrica*, 73(2), 417-458.

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.

SANDGREN, S., and STROM, B. (2005) "Peer effects in primary school: Evidence from age variation," Annual Conference EALE, Prague

SCHEIKE, T. H., and JENSEN, T. K. (1997): "A discrete survival model with random effects: An application to time to pregnancy," *Biometrics*, 53(1), 318-329.

SCHNEEWEIS, N., and WINTER-EBMER, R. (2007): "Peer effects in Austrian schools," *Empirical Economics*, 32(2-3), 387-409.

SHARP, C. (1995): "Whats Age Got to Do with It - a Study of Patterns of School Entry and the Impact of Season of Birth on School Attainment," *Educational Research*, 37(3), 251-265.

SHARP, C., HUTCHISON, D., and WHETTON, C. (1994): "How Do Season of Birth and Length of Schooling Affect Childrens Attainment at Key Stage-1," *Educational Research*, 36(2), 107-121.

SHIBLEY HYDE, J., FENNEMA, E., and LAMON, S. (1990): "Gender differences in mathematics performance: A meta-analysis," *Psychological Bulletin*, 107(2), 139-155.

SLAVIN, R. E. (1987): "Ability grouping and student achievement. Review of Educational Research," *Review of Educational Research*, 57(3), 293-336.

SMITH, A. (2004) "Making Mathematics Count The report of Professor Adrian Smith's Inquiry into Post-14 Mathematics Education," Department for Education and Skills, London

SMITHERS, A., and ROBINSON, P. (2006) "The Paradox of Single-Sex and Co-educational Schooling," Headmasters' and Headmistresses' Conference Centre for Education and Employment Research, University of Buckingham

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.

STOCK, J., and YOGO, M. (2005): "Testing for Weak Instruments in Linear IV Regression" in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by D. W. K. Andrews, and J. H. Stock. Cambridge: Cambridge University Press, 80–108.

STROM, B. (2004) "Student achievement and birthday effects," CESifo-Harvard University/PEPG Conference on "Schooling and Human Capital in the Global Economy: Revisiting the Equity-Efficiency Quandary", CESifo Conference Center, Munich

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.

URQUIOLA, M. (2005): "Does School Choice Lead to Sorting? Evidence from Tiebout Variation," *The American Economic Review*, 95(4), 1310-1326.

VIGDOR, J. L., and NECHYBA, T. J. (2007): "Peer Effects in North Carolina Public Schools" in *Schools and the Equal Opportunity Problem*, edited by L. Woessmann, and P. E. Peterson. Cambridge, Massachussetts: The MIT Press, 73-102.

WHITMORE, D. (2005): "Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment," *American Economic Review*, 95(2), 199-203.

WILSON, D. (2004): "Which ranking? The impact of a 'value-Added' measure of secondary school performance," *Public Money & Management*, 24(1), 37-45.

WINKLER, D. R. (1975): "Educational Achievement and School Peer Group Composition," *Journal of Human Resources*, 10(2), 189-204.

WINSTON, G., and ZIMMERMAN, D. J. (2003): "Peer effects in higher education" in *College Decisions: How students actually make them and* *how they could*, edited by C. Hoxby. Chicago: University of Chicago Press.

YOUNGER, M., and WARRINGTON, M. (1996): "Differential achievement of girls and boys at GCSE: Some observations from the perspective of one school," *British Journal of Sociology of Education*, 17(3), 299-313.

YOUNGER, M., WARRINGTON, M., GRAY, J., RUDDOCK, J., MCLELLAN, R., BEARNE, E., KERSHNER, R., and BRICHENO, P. (2005) "Raising Boys' Achievement," Research Report RR636, Department for Education and Skills, London

YOUNGER, M., WARRINGTON, M., and WILLIAMS, J. (1999): "The gender gap and classroom interactions: reality and rhetoric," *British Journal of Sociology of Education*, 20(3), 325-341.

ZIMMER, R. W., and TOMA, E. F. (2000): "Peer Effects in Private and Public Schools across Countries," *Journal of Policy Analysis and Management*, 19(1), 75-92.

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