The Good, The Bad and The Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools*

Victor Lavy‡, Olmo Silva♠ and Felix Weinhardt†

March 2009

Abstract

In this paper, we study the scale and nature of ability peer effects in secondary schools. In order to shed light on the nature of peer effects, we investigate which segments of the peer ability distribution drive the impact of average peer quality on students' achievements. Additionally, we study which quintiles of the pupil ability distribution are affected by different measures of peer quality. To investigate these issues, we use data for all secondary schools in England for four cohorts of age-14 (9^{th} grade) pupils sitting for their age-14 national tests in 2003/2004-2006/2007. We base our identification strategy on within-pupil regressions that exploit variation in achievements across the three compulsory subjects (English, Mathematics and Science) tested at age-14. Further, by focusing on pupils with little variation in prior achievements measured at the end of primary school, we identify a natural experiment that completely avoids biases due to endogenous sorting of pupils. We find significant and sizeable ability peer effects that mainly reflect the positive impact of the very academically bright peers and the negative impact of the very worst pupils. Moreover, we find some interesting and policy relevant heterogeneity along the dimensions of pupils' ability and gender. Finally, we show that our results are driven by peers' academic ability, and not related to their family background.

* We would thank the Center for Economic Performance at the London School of Economics for seed money for this project and Brendon McConnel for outstanding research assistantship. We would also like to thank Steve Gibbons and Steve Machin for helpful comments and discussions. All remaining errors remain our own.

‡ Hebrew University of Jerusalem, Royal Holloway University of London, CEP, CEPR and NBER. v.lavy@rhul.ac.uk or msvictor@huji.ac.il
♠ London School of Economics, CEP and IZA. O.Silva@lse.ac.uk
† London School of Economics and CEP. F.J.Weinhardt@lse.ac.uk

1. Introduction

The estimation of peer effects in the classroom and at school has received intense attention in recent years. Several studies have presented convincing evidence about race, gender and immigrants' peer effects[1], but important questions about the scale and nature (i.e. the 'origins') of ability peer effects in primary and secondary schools remains open, with little conclusive evidence.[2] In this paper we study ability peer effects in educational outcomes between schoolmates in secondary schools in England. We first investigate the size (i.e. the 'scale') of the effect of average peer quality on the outcomes of students, and then explore which segments of the ability distribution of peers drive the impact of average peer quality (i.e. the 'nature'). In particular, we study whether the extreme tails of the ability distribution of peers, namely the exceptionally low- and high-achievers, account for most or all of the effect of average peer quality on the educational outcomes of other pupils.

To do so, we use data for all secondary schools in England for four cohorts of age-14 (9^{th} grade) pupils entering secondary school in the academic years 2001/2002 to 2004/2005 and sitting for their age-14 national tests in 2003/2004-2006/2007. We link this information to data on pupils' prior achievement at age-11, when they sat for their end of primary education national tests, which we exploit to obtain pre-determined measures of peer ability in secondary schools. In particular, we construct measures of average peer quality based on pupils' age-11 achievements, as well as proxies for the very high- and very low-achievers, obtained by identifying pupils who are in the highest or lowest 5% of the (cohort-specific) national distribution of cognitive achievement at age-11. Note that the way we measure peer ability is a major improvement over previous studies. In fact, the vast majority of previous empirical evidence on ability peer effects in schools arises from studies that examine the effect of average background characteristics, such as parental schooling, race and ethnicity on students' outcomes (e.g. Hoxby, 2000 for the US, and Ammermueller and Pischke, 2006 for several European countries). A limitation of these studies is that they do not directly measure the academic ability of students' peers, but rely on socio-economic background characteristics as proxies for this. Additionally, our measures of peer quality are immune, for two reasons, to refection problems (Manski, 1993). First, we identify peers' quality based on pupils' test scores at the end of primary education, before students change school and move on to the secondary phase. As a consequence of the large reshuffling of pupils in England during this transition, on average secondary school students meet 87% 'new peers' at secondary schools, i.e. students that do not come from the same primary school. Secondly and crucially, we are able to track pupils during this

[1] Recent examples include Angrist and Lang (2004) on peer effects through racial integration; Hoxby (2000) and Lavy and Schlosser (2007) on gender peer effect; and Gould, Lavy and Paserman (forthcoming) on the effect of immigrants on native students.

[2] One exception is Sacerdote (2001), who presents evidence on ability peer effects in college based on co-residence of randomly paired roommates in university housing.

transition, which means that we can single out new peers from old peers, and construct peer quality measures separately for these two groups. In our analysis, we focus on the effect of new peers ability on pupil achievement (while controlling for old peers' quality), thus completely by-passing reflection problems.[3]

Our results show that having peers of high average ability significantly improves the cognitive performance of schoolmates, but this effect is primarily driven by peers who are at one of the two extreme ends of the ability distribution, namely the very 'good' and very 'bad' peers. More specifically, the effect of low ability peers is significantly negative, and the effect of high ability peers significantly positive, although the former is much larger and more precisely estimated. Interestingly, we find that the negative effect of very weak peers does not vary by the ability of regular students, whereas the positive effect of very bright peers is more positive for pupils below the median of the ability distribution, and it turns negative for pupils in the top 15% of the ability range. We further explore the heterogeneity of our findings, in particular along the lines of family background and gender. We find that our results are very similar for pupils who are eligible for free school meals (a proxy for family income) and for those who are not, and that the positive effect of 'good' peers and the negative effect of 'bad' ones are almost unaffected if we breakdown our measures of peers quality according to their free school meals eligibility. This is a very important as it suggests that our main findings are driven by peers' academic ability, and not related to their family background. Additionally, we find some interesting heterogeneity of peer effects by gender, especially regarding the impact of the very bright pupils in the class.

Besides providing some novel insights about the nature of ability peer effects, our paper presents a new identification approach that allows us to improve on the (non-experimental) literature in the field and to identify the effects of peers' ability while avoiding biases due to endogenous selection and sorting of pupils, or omitted variables issues. Indeed, the distribution of pupils' characteristics in secondary schools in England, like in many other countries, reflects a high degree of sorting and selection by ability. For example, using pupils' age-11 nationally standardized test scores as an indicator of ability, we find that the average ability of peers and the pupil's own ability in secondary school are highly correlated. This is despite the fact that most students change school when moving from primary to secondary education, and that on average pupils meet 87% 'new peers'. Similarly, there is a high correlation between pupils' and their peers' socioeconomic background characteristics, which is further evidence that students are not randomly assigned to secondary schools and that the very top and very low achievers are typically clustered in high- and low-achieving schools.

In order to overcome this selection problem, we rely on within-pupil regressions that exploit variation in achievements across the three compulsory subjects (English, Mathematics and Science) tested

[3] Note that Gibbons and Telhaj (2008) exploit a similar intuition.

at age-14. We further exploit the fact that students were tested on the same three subjects at age-11 (at the end of primary schools), so that we can measure peers' ability separately by subject. However, as we shall see below, sorting into high-school is still evident in the correlation between the within-student across-subject variation in age-11 achievements, and the variation in peers' ability across subjects. To overcome this 'sophisticated' form of selection, we propose a natural experiment that takes advantage of a sub-sample of pupils that faces large deviations in peers' ability across subjects, but no correlation between this variation and the within-pupil own variation in ability across subjects. These two desired features are simply achieved by focusing on students who have practically no differences in their primary school test scores across the various subjects. By definition, in this 'experimental' sample there is no correlation between the within-pupil across-subject variation in ability and the variation in peers' ability. Furthermore, our results are largely unchanged irrespective of whether we directly control for pupil subject-specific ability or not (i.e. own age-11 attainment). Based on this natural experiment we document significant effects of secondary school peers' ability on pupils' own test scores at age-14. We are also able to enhance the credibility of our identification strategy by contrasting the estimated treatment effects to those based on two 'placebo' treatments, obtained by switching peer quality measures across subjects. [4]

Remarkably, the causal relationships that we estimate between pupil performance in a specific subject and peer quality in that same subject in our 'experimental' sub-sample have a very similar size and precision if we go on to consider pupils with larger across-subject variation in their prior achievements, and if we allow for a significant degree of correlation between the within-student across-subject deviations in age-11 test scores and the variation in peers' quality across subjects. Furthermore, our results are largely unchanged irrespective of whether we directly control for pupil subject-specific ability or not (i.e. own age-11 attainment). Thus, we conclude that the observed correlation between the within-pupil heterogeneity across subjects and the variation of schoolmates' quality across subjects does not confound the treatment effect of peers' ability once the estimation is based on within-student variation. This finding is very important as it suggests that the estimates based on our 'experimental' sample have strong external validity for secondary students at large in England.

The rest of the paper is organized as follows. The next section reviews the recent literature on peer effects, while Section 3 describes the identification strategy. Section 4 discusses the data and the construction of our samples, while Section 5 assesses the validity of our identification strategy. Next, Section 6 reports our main OLS and within-student estimates of ability peer effects and Section 7 presents some heterogeneity in our findings. Finally, Section 8 provides some concluding remarks.

[4] A number of recent studies have also used explicit random or quasi-random assignment to classes or schools, or other natural experiments, for example, Sacerdote (2001), Zimmerman (2003), Angrist and Lang (2004), Arcidiacono and Nicholson (2005), Hanushek et al. (2003) and Gould, Lavy and Paserman (forthcoming).

2. Related literature

For a long time social scientists have been interested in understanding and measuring the effects of peers' behavior and characteristics on individual outcomes, both empirically (e.g. Coleman, 1966) and theoretically (e.g. Becker. 1974). The basic idea is that group actions or attributes might influence individual decisions and outcomes, such as educational attainment. Despite its intuitiveness, the estimation of peer effects is fraught with difficulties, and many of the related identification issues have yet to find a definitive answer. In particular, Manski (1993) highlights the perils of endogenous group selection and the difficulty of distinguishing between 'contextual' and 'endogenous' peer effects. In practice, most studies have ignored this last distinction and focused on reduced form estimation as outlined by Moffit (2001), where peer group characteristics are used to explain differences in individual outcomes. Even then, the literature has attempted to by-pass those biases that arise because of endogenous sorting or omitted variables, and has not yet reached a consensus regarding the size and importance even of these reduced form effects.

In particular, two main issues have taxed researchers interested in the identification of the causal effect of peer quality in education. Firstly, it is widely recognized that one pupil's peer group is evidently self-selected and hence the quality of one's peers is not exogenous to one pupil's own quality and characteristics. In fact, there is a well established literature on the link between school quality and house prices (Black, 1999, Gibbons et al., 2009 and Kane et al., 2006), suggesting that pupils are segregated into different neighborhoods and schools by socio-economic status. Failing to control for all observable and often unobservable factors that determine individual sorting and their achievement would result in biased estimates of ability peer effects. Secondly, peer effects work in both directions, so that peer achievements are endogenous to one pupils' own quality if students have been together for a while. This mechanical issue, known as the 'reflection problem', is particularly difficult to undo unless the researcher is able to reshuffle group formation and belonging, and measure peers' quality in ways that are predetermined to interactions within the group.

To account for these difficulties, recent years have seen a variety of identification strategies. Different studies have exploited random group assignments (Sacerdote, 2001; Zimmerman, 2003; Duflo et al., 2008), within-school random variation (Hoxby, 2000; Hanushek et al., 2003; Ammermualler and Pischke; 2006), instrumental variables (Goux and Maurin, 2007) or sub-group re-assignments (Sanbonmatsu et al., 2004).[5] Only recently, Lavy and Schlosser (2007), Lavy et al. (2008) and Duflo et al. (2008) have tried to enter the 'black box' of ability peer effects in Israel and Kenya, respectively, and have explicitly focused on understanding the mechanisms through which social interactions could

[5] Other examples include: Aizer (2008), Bayer et al. (2004), Bifulco et al. (2008), Burke and Sass (2004), Carrell and Hoekstra (2008), Figlio (2007), Lefgren (2004), Nechyba and Vidgor (2005) and Vidgor and Nechyba (2004).

operate. In particular, Duflo et al. (2008) exploits random assignment of pupils to classes as way to identify peer effects. The authors focus on pupils in Kenyan primary schools with a single first year class (and average class size of over eighty), which is split in half as an additional teacher is assigned to each school. New classes are either formed based on previous results of the pupils (tracking) or randomly, and the assignment of pupils to either of these 'treatments' is also random. The authors find significant improvements from ability-tracking in primary schools and attribute this result to the fact that more homogeneous groups of students might be taught more effectively. (Lavy et al., 2008) present related evidence of significant and negative effect of a high fraction of low ability students in the class (repeaters) on the outcomes of other pupils, which might arise through classroom disruption and decrease in attention paid by the teacher. However, it has to be noted that the primary school setting in Kenya might not be fully comparable to more developed countries.

The study that is closest to ours in terms of context and data is Gibbons and Telhaj (2008) who also estimate peer effects for eleven to fourteen year old pupils in English secondary schools, and with whom we also share the focus on the re-shuffling of peers that is caused by the primary to secondary school transition. While this presents an effective way to account for the simultaneity and reflection problems, the study by Gibbons and Telhaj (2008) attempts to control for the sorting of pupils with similar abilities and background in the same secondary schools by allowing for primary and secondary school fixed-effect interactions and trends. As it turns out, this does not fully eliminate the correlation between pupils' own ability and peer quality, as measured by students' end of primary school achievements. Their results provide little evidence of sizeable and significant peer effects in their linear-in-means specifications.

To our knowledge, our study is the first one to rely on pupil fixed effects and inter subject differences in achievement to address identification issues of peer effects.[6] Additionally, as already mentioned, by focusing on pupils with little variation in prior achievements measured at the end of primary school, we further identify a 'natural experiment' that completely avoids biases due to endogenous selection and sorting of pupils. This allows us to achieve a clean identification of the causal effect of peer quality. In the next section we spell out in more details our empirical strategy.

3. Empirical Strategy

3.1. Identification Strategy

The main problem with identifying the effect of the ability composition of peers on pupil educational achievements is that peer quality measures are usually confounded by the effects of unobserved correlated

---

[6] Lavy (2009) uses the same approach and examines subject-to-subject variation in outcomes for the same student to investigate the effect of instructional time on pupil academic achievements.

factors that affect students' outcomes. This correlation could arise if there is selection and sorting of students across schools based on ability differences, or if there is a relation between the average students' ability in one school and other characteristics of that school (not fully observed) that might affect students' outcomes. The approach commonly used in several recent studies relies on within-school variations in the ability distribution of students across adjacent cohorts or across different classes (e.g. Hoxby, 2000; Gould et al., 2004; Ammermualler and Pischke, 2006; Lavy and Schlosser, 2007; and Lavy et al., 2008). This method potentially avoids both sources of confounding factors highlighted here above, although the identifying assumption is that the variation of peer quality over time (or across classes) is purely idiosyncratic and uncorrelated with students' potential outcomes and background.

In this paper, we suggest an alternative approach for overcoming the potential selection/sorting and omitted variable biases, namely we examine subject-to-subject variation in outcomes for the same student and investigate if they are systematically associated with the subject-to-subject variation in peer ability. Thus, the ability peer effect that we study here is subject specific. Stated differently, in this paper we question whether pupils who have school peers that have on average higher ability in subject *j* (e.g. Mathematics) than in subject *i* (e.g. Science), have better cognitive performance in subject *j* than in subject *i*. More formally, using test scores in multiple subjects, we estimate the following pupil fixed-effect equation:

$$A_{iqs} = \alpha_i + \beta_q + \gamma_s + \delta \ P_{qs} + \varepsilon_{iqs} \qquad (1)$$

where *i* denotes pupils, *q* denotes subjects and *s* denotes schools. A_{iqs} is an achievement measure for student *i* in subject *q* at school *s*. In our analysis, we focus on test scores in the three compulsory subjects (English, Mathematics and Science) assessed at age-14 during the national tests; these are denoted in England as Key Stage 3 (KS3; more details are presented in Section 4). Additionally, α_i is a student fixed effect, β_q is a subject specific effect, and γ_s is a school effect. Further, P_{qs} captures the average ability of peers in subject *q* of peers in secondary school *s* as measured by the test scores in a given subject in the national tests taken by students at age-11 at the end of primary school. These are denoted as Key Stage 2 (KS2). Finally, ε_{iqs} is an error term, which is composed of a pupil-specific random element that allows for any type of correlation within observations of the same student and of the same school. The coefficient of interest is δ which captures the effect of having higher or lower ability peers on students' achievement.

A summary of the ideas at the base of our empirical strategy is that we compare the test scores of a student in three different subjects to the average ability of peers in each of these subjects. Of course, for a given pupil, each of these subjects 'faces' the same pupil characteristics and the same school environment. In a nutshell, pupil fixed-effects (capturing his/her average ability across subjects) and school fixed effects are 'absorbed' and taken care of in our specification. Nevertheless, this setup does not preclude the possibility that selection and sorting of students also is partly based on subject specific abilities of pupils and peers.

To completely rule out this possibility, we rely on the following 'natural experiment': we identify a sample of pupils for which there is little or no differences across KS2 test scores in the three subjects, but still face substantial variation in the quality of secondary school peers across the three different subjects. In this sub-sample, there can not be any correlation between pupil subject-specific ability and subject-specific peers' ability. Therefore any correlation between the KS3 test scores of pupils in secondary schools and the ability of their peers (based on their KS2) identifies the causal effect of peer quality on educational outcomes. Importantly, in our analysis we will compare estimates based on this 'experimental' sample to the results that we obtain from samples that exhibit some degree of correlation between the within-pupil and the within-peers variation in subject-specific abilities. One remarkable finding is that our estimates of the ability peer effects do not change as we stretch our sample to include pupils in schools that display some subject-specific sorting. In other words, our strategy that controls for a student fixed-effect eliminates the sorting/selection and omitted variable biases even in samples that allow for high within-pupil across-subject variation in KS2 test scores.

As discussed above, we are also interested in finding out which segments of the peer ability distribution are driving the mean ability peer effect that we document. Therefore, we also estimate models where we add treatment variables that measure the proportion of peers who are very 'good' or very 'bad'. To do so, we choose the top and bottom 5% in the (cohort-specific) national distribution of KS2 test scores as the cut off point to determine the very high-ability (P_{qs}^{h}) and the very low-ability (P_{qs}^{l}) pupils (more details in the data section). We then estimate the following equation:

$$A_{iqs} = \alpha_i + \beta_q + \gamma_s + \delta_1 P_{qs} + \delta_2 P_{qs}^{h} + \delta_3 P_{qs}^{l} + \varepsilon_{iqs} \qquad (2)$$

where δ_2 measures the effect of the proportion of peers who are in the top 5% of the national distribution of KS2 test scores, and δ_3 captures the effect of the proportion of pupils who are in the bottom 5%. Lastly, the parameter δ_1 measures the effect of the average ability of all other peers.

Note that since we use four cohorts of data, we add a time subscripts to all variables in the equation and estimate the following equation based on pooled cross sections:

$$A_{iqst} = \alpha_i + \beta_q + \gamma_s + \delta_1 P_{qst} + \delta P_{qst}^{h} + \delta P_{qst}^{l} + \varepsilon_{iqst} \qquad (3)$$

3.2. How We Measure Peers Ability

A key requirement for our empirical approach is that our proxies of peer ability are based on a pre-determined measures of students' ability that have not been affected by the ability of his/her peers and thus do not suffer from reflection problems. As already discussed, the longitudinal structure of the administrative data that we use allows us to link peers' KS2 test scores taken at the end of primary school (6^{th} grade) to students' KS3 achievements three years later, that is 9^{th} grade in secondary school. Additionally, by following individuals over time we are able to identify which secondary school students come from the same primary, and that is we can identify who the 'new peers' and the 'old peers are. On average, about 87% of pupil *i*'s peers in secondary school did not attend the same primary institution as student *i*, and therefore their KS2 test scores could not be affected by this pupil. Thus, in our analysis, we construct peer quality measures separately for 'new peers' and 'old peers', and focus on the effect of former on pupil achievement to avoid reflection problems. Note also that in most of our empirical work we include measures of the quality of old peers as additional controls. However our estimates are not sensitive at all to the inclusion of these variables because they are practically orthogonal to the measures of the ability of new peers (conditional on pupil fixed effects).

Before moving on two important remarks are worth being made. First, we use information about the school that a pupil is attending at age-12 (7^{th} grade), when he/she enters secondary education, to define our base population; similarly our three measures of peer quality 'treatment' (the good, the bad and the average peer quality) are also based on 7th-grade enrollment. This is because any later change in these proxies, for example as recorded at KS3, might be endogenous. Second, in implementing this methodology, we use the peers' ability measured at the grade and *not* at the class level because our data does not include class identifiers. However, even if this information had been available to us, we would have measured peer ability as described above because class placement might be endogenous, as parents and school authorities may have some discretion in placing students in different classes within a grade. This is not a very restrictive compromise because in other studies it has been shown that within a given school the average peer ability in a grade is highly correlated with the average peer ability in a class, and we have no reason to believe it is different in England.

To sum up, in order to implement our strategy we need to gather detailed information about the school where pupils completed their primary education and about the institution at which they start secondary schooling. Additionally, we need to link this information to pupil achievements across different subjects both at the end of primary school (KS2) and in 9^{th} grade (KS3). In the next section, we describe how the English school and pupil census allow us to gather all this information. First, however, we provide additional details about the English educational system.

4. Data and Descriptive Statistics

4.1. Institutional Background and Data Construction

Compulsory education in England is organized into five stages referred to as Key Stages. In the primary phase, pupils enter school at age 4-5 in the Foundation Stage then move on to Key Stage 1, spanning ages 5-6 and 6-7 (these would correspond to the 1st and 2nd grade in other educational system, e.g. in the US). At age 7-8 pupils move to Key Stage 2, sometimes – but not usually – with a change of school. At the end of Key Stage 2, when they are 10-11 (6th grade), children leave the primary phase and go on to secondary school where they progress through Key Stage 3 and 4. Importantly, the vast majority of pupils changes schools on transition from the primary to the secondary stage. Although a few schools still operate a Middle School system, bridging the primary and secondary phases, these are not considered here. At the end of each Key Stage, generally in May, pupils are assessed on the basis of standard national tests (SATS), and progress through the phases is measured in terms of Key Stage Levels, ranging between W (working towards Level 1) up to Level 5+ during primary education and Level 7 at KS3. Importantly for our research, at both KS2 (6th grade) and KS3 (9th grade) students are tested in three core subjects, namely Mathematics, Science and English, and their attainment are recorded in terms of the raw test scores, spanning the range 0-100, from which the Key Stage Levels discussed here above are derived.

The UK's Department for Children, Schools and Families (DCSF) collects a variety of data on all pupils and all schools in state education[7]. This is because the pupil assessment system is used to publish school performance tables and because information on pupil numbers and pupil and school characteristics is necessary for administrative purposes – in particular to determine funding. In fact, starting from 1996, a database exists holding information on each pupil's assessment record in the Key Stage SATS described above, throughout their school career. Additionally, starting from 2002, the DCSF has also carried out the Pupil Level Annual Census (PLASC), which records information on pupil's gender, age, ethnicity, language skills, any special educational needs or disabilities, entitlement to free school meals and various other pieces of information, including the identity of the school attended during years other than those when pupils sit for their Key Stage SATS. The PLASC is integrated with the pupil's assessment records (described above) in the National Pupil Database (NPD), giving a large and detailed dataset on pupil characteristics, along with their test histories. Furthermore, various other data sources can be merged in at school level using the DCSF Edubase and Annual School Census, which contain details on school institutional characteristics (e.g. religious affiliation), demographics of the enrolled students (e.g. fractions of pupils eligible for free school meals) and size (e.g. number of pupils on roll and number of teachers at the school).

[7] The private sector has a market share of about 6-7%. However, very little consistent information exist for pupils and schools in the private domain. For this reason, we do not consider private schooling in our analysis.

The length of the time series in the data means that it is possible for us to follow the academic careers of four cohorts of children through from age-11 (6th grade) to age-14 (9th grade), and to join this information to the PLASC data for every year of secondary schooling (7th to 9th grade). More in details, from our large and complex initial dataset, we are able to extract information for pupils who finished their primary education in the academic years 2000/2001 to 2003/2004, entered their secondary schools in the academic years 2001/2002 to 2004/2005, and sat for their KS3 exams in the academic years 2003/2004 to 2006/2007. Importantly, we use information on these four cohorts as the core dataset in this study because this only time window where we can identify the secondary school were pupils *start* their secondary education, and not only the one where they sit for their KS3 national tests. As explained above, this is crucial to our analysis because we want to be able to measure peer exposure at the beginning of secondary schooling (in 7th grade), and not after two year, at KS3. Our concern here is that students and their parents might react to peer quality observed during the first two years of secondary education by changing school. This would imply that peer quality measures identified at any later stage are endogenous. Additionally, as highlighted above, for every pupil we have information about the primary school where he/she took the KS2 exams, which implies that we are able to single-out those secondary school peers that are 'new peers' from those who instead came from the same primary school (which we label 'old peers').

Using this set of information we construct a variety of peer quality measures based on pupil achievements at KS2 in the three core subjects. In order to do so, we start by using the KS2 test scores, separately by subject and cohort, to assign each pupil to a percentile in the cohort-specific national distribution of KS2 tests for that subject. We then go on to create three separate measures of peer quality. First of all, and following the vast majority of studies in the field, we compute the average attainments of peers in the grade (excluding the pupil under consideration). Next, we go on to create two measures that are meant to capture peer effects coming from very bright and very worst students in the school, namely: the fraction of pupils below the 5th percentile or above the 95th percentile of the cohort-specific national distribution of Key Stage 2 tests (by subject). Note again that we are able to construct our peer quality measures separately for 'new peers' and 'old peers'.

Before moving on to some descriptive statistics and our core analysis, it is important to discuss some restrictions that we have imposed on our data in order to obtain a balanced panel of pupil information in a balanced panel of schools with stable characteristics. First of all, we have selected only pupils with valid information on their KS2 and KS3 tests for whom we can also match individual background characteristics and the identity of the school where they start their secondary education using PLSAC. Given the quality of our data, this implies that we drop only less than 2.5% of our initial data. Next, we have focused on schools that are open in every year of our analysis, and have further dropped secondary schools that have a year-on-year change of entry-cohort size of more than 75% or enrolments below 15 pupils. While the former restriction excludes schools that were exposed to large shocks that might confound our analysis, the latter excludes schools that are either extremely small or had many missing observations. These restrictions imply that we loose less than 2.5% of our observations. We have also excluded selective schools (e.g. Grammar schools) from our analysis, as these schools can actively choose their pupils based on their ability (about 8% of our original sample).

Furthermore, we apply some further restriction based on two of our peer measures, namely the fraction of bottom 5% and top 5% pupils in one student's cohort, in order to exclude schools with particularly high or persistently low shares of 'good' and 'bad' peers. In more details, we drop schools where the fractions of pupils below the 5^{th} percentile or above the 95^{th} percentile of the cohort-specific KS2 national distribution exceeds 20%, and schools that do not have any variation over the four years in the fractions of pupils below the 5^{th} or above the 95^{th} percentile. Importantly, this last restriction predominantly trims schools that have no students in either the top 5% or bottom 5% of the ability distribution in any year, and thus would not contribute to the identification of peer effects. The two combined restrictions imply that we drop an additional 10% of our sample. Since this seems a large share, we checked that our main results are not affected when we omit these restrictions.

In conclusion, our final dataset includes a balanced panel of more than 1,500,000 pupils for whom we can observe complete information in terms of KS2 and KS3 test scores, individual and family background characteristics, and both primary and secondary school level information from age-11 to age-14. In the next section, we present some descriptive statistics for various samples.

4.2 Some Descriptive Statistics

In Tables 1 we present some descriptive statistics for both the full sample and for our 'natural experiment' sample of pupils with little or no variation in their KS2 performance across subjects. More precisely, the restricted sample only considers pupils for whom the standard deviation in their KS2 percentile ranking across the three subjects is at most 3. This roughly includes 6% of our full sample. Further note that the we present descriptive statistics separately for pupils in the central part of the ability distribution, that is pupils above the 5^{th} percentile and below the 95^{th} percentile of KS2 test score distribution, and students in either the top 5% or bottom 5% tails of the ability distribution, that is our 'good' and 'bad' peers.

In the top panel of the Table we describe pupils' test scores at KS2 and KS3. Unsurprisingly, the first column shows that for the full sample test score percentiles are centered just below 50, for all subjects and at both Key Stages.[8] In the second column, we focus instead on our 'natural experiment' pupils with little or no variation in their KS2 achievement across subjects. Compared to the full sample,

[8] They are not exactly equal to 50 because slightly more than one percent of pupils scored zero points in each of the subjects and hence all belong to the first percentile.

this group achieve on average about 7 percentiles worse in every individual subject. This reflects the fact that while many good pupils are equally bright across the three subjects, more pupils tend to be equally bad across English, Mathematics and Science. Finally, the last two columns illustrates how pupils with at least one subject in either the top 5% or the bottom 5% of the ability distributions score at their KS2 and KS3 tests. By construction, pupils in top 5% of the KS2 test score distribution perform much better than any other pupil in their KS2 exams, while the opposite is true for pupils in the bottom 5% tail. More interestingly, this stark ranking is not changed when we look at KS3 test scores, for all subjects, with little evidence of mean reversion in achievements of pupils at the very top and very bottom between age-11 and age-14. To further substantiate this point, in Appendix Table 1 we analyze more thoroughly the KS3 percentile ranking of pupils in the top 5% and bottom 5% of the KS2 achievement distribution. The table shows that, for all subjects, about 80% of the pupils ranking in the bottom 5% at KS2, rank in the bottom 20% of the KS3 distribution, and between 60% and 70% of them are in the bottom 10%. At the opposite extreme, around 80% of pupils ranking in the top 5% at KS2 remains in the 20% of the KS3 achievement distribution, with the vast majority still scoring in the top 10%. This reinforces the idea that our 'good' peers and 'bad' peers are consistently amongst the brightest and worst performers.

The second panel of Table 1 presents more information on pupil background characteristics. Here our selected 'natural experiment' sample is virtually identical to the full sample of pupils. On the other hand, pupils with at least one subject in the bottom 5% are less likely to have English as their first language, are less likely to be of White British ethnic origins and are more likely to be eligible for free school meals (FSM, a proxy for family income), while the opposite is true for pupils with at least on subject in the top 5%. However, the differences in family background are much less evident than those in terms of academic ability presented in Panel A. This suggests that peer ability measures defined in terms of pupil background might severely underestimate differences in peers' academic quality.

Finally, in Panel C we report school characteristics for the various sub-groups. The average cohort size at the start of secondary school in 7^{th} grade is approximately 200, and around two thirds of all pupils attend Community schools, while about 15% of the pupils attend a religiously affiliated state-school. Furthermore, pupils with at least one subject in the top 5% of the ability distribution are less likely to attend a Community school and more likely to be in a faith school than pupils in the central part of the ability distribution and pupils with at least one subject in the bottom 5%. However, these differences are not remarkable.

In Table 2, we move on to present some descriptive statistics of our treatments, both for the full sample and for the 'natural experiment' sample of pupils. Statistics are presented separately for all peers and for new peers only. Note once again that, as reported at the bottom of the table, on average pupils face 87% new schoolmates, although the distribution of new peers is highly right-skewed, with many more pupils facing 100% new schoolmates than zero. Panel A summarizes the average peer quality, computed as the average KS2 percentile rank of peers in a given subject (excluding the pupil under consideration). Moving across the columns reveals that average peer quality is not dissimilar in the full sample and in our 'natural experiment' sample. This is very reassuring as it shows that our restriction on the variation of pupils' KS2 achievement across subjects does not affect the quality (and the variation) of the peers that the students are confronted with.

In Panel B and Panel C, we present descriptive statistics for our proxies for 'good' and 'bad' peers. Once again, we present this separately for all peers and new peers. By construction, the fractions of top 5% and bottom 5% 'new peers' in the incoming cohort are smaller than the respective fractions including all peers. However, the descriptive statistics for our experimental sample are virtually identical to those for the full sample. Once more, this suggests that students in our 'natural experiment' sample are exposed to peers of a similar quality as pupils in the full sample.

5. Evidence on the Validity of the Identification Strategy

We start our analysis by presenting some evidence on the validity of our identification strategy. To do so, we test whether the within-secondary school variation in peers' average KS2 test scores across the three subjects is associated with within-pupil variation in KS2 test scores. Figure 1 plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained from regressions of pupil KS2 achievements (percentiles) on the average achievement of new peers at KS2. All regressions include pupil fixed effects and control for old peer quality. To obtain the figure, 18 different regressions were estimated over different cumulative bands of the standard deviation of KS2 attainments across subjects; these spanned the interval s.d.≤11.5 to s.d.≤3, in steps of 0.5. The sample defined by a standard deviation of at most 3 is our 'natural experiment' sample, which includes roughly 6% of the sample. At the opposite extreme, the sample with s.d.≤11.5 includes about 50% of the full sample, while the set of pupils with s.d.≤7.5 includes about 25% of the full sample. Additional details can be found in the note to the figure.

Figure 1 shows that in all samples but the one defined by s.d.≤3 there is significant positive correlation between one pupil's own KS2 achievements and the average achievement of his/her new peers in secondary school at KS2. This is clear evidence of a significant degree of positive selection and sorting into high school even when a student fixed effect is included in the regressions. However, the magnitude of this imbalance is dramatically diminished as we reduce the within-pupil variation in KS2 test scores, and in our natural experiment sample with s.d.≤3 this correlation is no longer significantly different from zero, although it is still positive but very small.

In Table 3, Column 1, we provide more evidence on the degree of balancing that is achieved in the sample of s.d.≤3. Starting from Panel A, OLS estimates (i.e., without student fixed effect) show a strong positive correlation between student characteristics and the KS2 achievements of new peers. Equally strong is the OLS relationship between one student's own KS2 achievements and the KS2 achievements of his/her new peers. However, as we add a student fixed-effect to the regression, these imbalances are eliminated. Obviously, the estimates for student characteristics (which do not vary by subject; e.g. free school meals eligibility) are exactly zero, as implied by the nature of the within-pupil regression. But adding the pupil fixed effect also eliminates the strong and significant positive correlation of pupils' and their peers' KS2 achievements. Indeed, the OLS estimate of this link is 0.291 whereas the within student estimate is 0.002. Of course, we can reduce the latter estimate to exactly zero by reducing standard deviation of pupil KS2 achievement to zero, but that will be at the expense of a further reduction in the sample size. More interestingly, however, as we will show below, our estimated peer effects will be almost identical as we increase the standard deviation to up to 11.5.

Next, we assess the balancing of pupils' characteristics with respect to our two other treatments, that is the fraction of 'good' and 'bad' peers. Figure 2 plots regression coefficients and confidence intervals obtained from regressing pupil KS2 achievements (percentiles) on the percentage of top 5% pupils (new peers) and percentage of the bottom 5% pupils (new peers) in the 7^{th} grade cohort. As for the estimates presented in Figure 1, here too all regressions include pupil fixed-effects and control for old peer quality; similarly, 18 different regressions were estimated over cumulative bands of the standard deviation of the KS2 test scores across subjects to obtain the various point estimates and confidence intervals. Figure 2 shows remarkably symmetric convergence from high to zero correlation between each of the two treatment measures and the KS2 achievements of pupils as we reduce the within student standard deviation in KS2 test scores. The estimate of the effect of the percentage of top 5% new peers is positive and significant, but it converges to almost zero and becomes insignificant at s.d. $\leq$3. Symmetrically, the estimate of the effect of the percentage of the bottom 5% new peers is negative and significant, until it converges to an insignificant and small negative value at s.d.$\leq$3. These patterns of selection and sorting are of course a reflection of the same pattern that we saw in Figure 1 for the balancing of the average quality of peers.

To further illustrate the validity of our strategy, Columns 2 and 3 of Table 3 present OLS and within-pupil evidence on balancing of individual characteristics with respect to the top and bottom peer quality treatments for the 'natural experiment' sample. Note that the estimates of the effects of the top 5% peers and bottom 5% peers are obtained from one single regression including both treatments (and controls for old peer quality). Similarly to the evidence presented in Column 1, OLS estimates reveal large positive selection with respect to the top 5% new peers and strong negative selection relative to the bottom 5% new peers. However, the within-pupil regression results suggest that this selection is entirely eliminated once controlling for pupil fixed-effects.

Overall, by focusing on within-pupil across-subjects variation and concentrating on the sub-sample of students with little variation in KS2 attainment, we are able to eliminate all the observed associations between peer ability and lagged test scores (and pupil, family, school and any other fixed characteristics) of secondary school pupils. However, as already noticed, the sample that yields this 'experimental' quality of perfect balancing is not representative of the full sample of secondary pupils in England, as it includes a relatively large proportion of low KS2 achievers and a relatively high proportion of KS2 achievers (although the former is bigger than the latter). Therefore, we will not be able to extrapolate the evidence based on our 'natural experimental' sample to the whole population. Nevertheless, as already mentioned, the evidence from this 'natural experiment' sample is virtually identical to the evidence based on larger samples, and in particular on the set pupils with a s.d.≤11.5. As this sample is very similar in its mean characteristics to the full sample of pupils (see Appendix Table 2), there is little reason to be concerned about the external validity of our results, which potentially apply to the whole population of secondary school pupils in England.

6. Results

6.1. Effects of Peer Ability in the Natural Experiment Sample

We begin our discussion of the results by presenting estimates of the impact of the peer quality on pupil outcomes at KS3 obtained using our 'natural experiment' sample (s.d. of KS2 ≤3). These are reported in Table 4. Columns 1 and 2 present OLS and within-pupil estimates of the effect of average peer quality. Next, Columns 3 and 4 present OLS and within-pupil estimates of the effect of the percentage of bottom 5% peers, while Columns 5 and 6 present estimates of the effect of the percentage of top 5% peers. Note that the estimates presented in the top panel are based on two different specifications. The first row reports estimates where only the average quality of previous peers (old peers) is included as a control, whereas the second row further includes one pupil's own KS2 attainments.

The estimates in Column 1 and 2 suggest that average peer quality has a positive effect on students' KS3 achievements. However, the within-pupil estimate is about a tenth of the OLS estimate, at 0.046 versus 0.444. Additionally, while the OLS estimate is highly sensitive to controlling for KS2, the within-pupil estimate is virtually identical in both specifications (0.046 and 0.045) and the estimated standard error is similarly unaffected. This is not surprising and it is a direct consequence of the perfect balancing in this sub-sample of treatment with respect to one pupil's own KS2 test scores. As for the size of the estimated peer effect, a 10 percentile increase in the average quality of peers would increase pupil own KS3 test scores by 0.46 of a percentile. This change amounts to around 0.02 of a standard deviation of the KS3 test score distribution. Further, the estimated effect implies that if a pupil is moved from the worse peer environment we observe in the data (where the average peer achievement in English, Mathematics or Science is at about the 25th percentile) to the best peer environment observed in the data (where the average peer achievement in one of the three subjects is around the 75th percentile), his/her score will rise by 2.3 percentiles in the national distribution. This gain amounts to 0.08 of the standard deviation of the KS3 score distribution. Since our results are estimated from regressions that include pupil fixed-effects, it is also instructive to understand how sizeable they are once we rescale them by the within-pupil variation in peer quality and KS3 attainments. In this case, we find that a 10 percentile increase in average peer quality (about four standard deviations in the within-pupil average peer quality distribution) amounts to around 0.06 of a standard deviation in the within-pupil KS3 distribution.

Is this effect large or small? To answer this question we need to compare our results to the effect of other inputs or interventions in secondary schooling but such evidence, especially with well identified treatment effects, is limited. Lavy (2009) estimates the effect of instructional time in secondary schools using the PISA 2006 data and reports an effect for OECD countries of 0.06 of a standard deviation (of the test score distribution) for an additional hour of classroom instruction, or 0.15 when considering the within pupil standard deviation of test scores across subjects. Estimates based on data from the UK yield a higher effect, 0.05 of a standard deviation for one more hour of school instruction, or 0.20 when focusing on within-pupil variation. These estimates imply that the ability peer effects that we estimate here is equivalent to the effect of two more hours of instruction in school. Another possible comparison is to the effect size of the peer effects estimated in Ammermueller and Pischke (2007) across classes within schools in six European countries. This study reports that one standard deviation change in their student background measure of peer composition leads to a 0.17 standard deviation change in reading test scores of fourth graders. Finally, Gibbons and Telhaj (2009) show that one stand deviation change in average peer quality is linked to about 0.05 of a standard deviation in the test score distribution of nine graders, although this estimate is still positively biased by endogenous sorting. All in all, in comparison to studies that focus on other school inputs and interventions, our estimates capture a medium sized effect.

In the next four columns of Table 4 we study whether 'good' peers and 'bad' peers similarly exert an effect on pupils' attainments at KS3. More precisely, the estimates in Column 3 and 4 present the effect of the percentage of the bottom 5% new peers, while the estimates in Columns 5 and -6 report the effect of the percentage of the top 5% new peers. Notice that these coefficients are obtained from one single regression including both treatments, and controlling for the quality of old peers. These results clearly show that having more students from the bottom 5% at school harms the academic performance of regular students, while sharing the school environment with more students from the top 5% improves the academic performance of regular students. In both cases the estimated effects are not sensitive to the inclusion of controls for lagged test scores (KS2). As for Columns 1 and 2, the OLS estimates are much larger than the within-pupil estimates, reflecting their large selection bias. Additionally, the peer effect estimate is much larger for the bottom 5% treatment than for the top 5% peer measure, at -0.148 versus 0.071. The magnitude of the effect of these two treatments can be assessed by scaling it according to the minimum and maximum values of the bottom and top treatment variables observed in the data. The minimum percentage is 0 for both treatments and the maximum percentage observed in the data is about 20% in both cases. Therefore a pupil who moves from 0 to 20% for the bottom quality peer environment will suffer a decline of KS3 test score of 3 percentiles, which amounts to 0.12 of the standard deviation of KS3 test score, or 0.33 if we consider the standard deviation of the within-pupil KS3 distribution. On the other hand, improving the peer environment from 0 to 20% in top quality peers will cause an increase of about 1.4 percentiles in KS3 test score, implying a change of 0.05 of a standard deviation or 0.18 of the standard deviation the within-pupil KS3 distribution. Note that these are rather sizeable changes, as they correspond to about 20 standard deviation changes in the within-pupil peer quality distribution, both for the top 5% and bottom 5% peers. More modest changes of around four standard deviations in the within-pupil distribution of 'good' and 'bad' peers (comparable to those used for the average peer quality here above), would respectively imply an increase of about 6% and a decrease of 16% of a standard deviation in the within-pupil KS3 distribution.

For these estimates too we would like to gauge an assessment of their relative size compared to other studies. To do so, suppose that our regular students are exposed to the following two treatments simultaneously: a reduction in the percentage of top5% and bottom 5% new peers from 5% each to zero percent. This change can be viewed as a move towards class homogeneity in terms of ability, that is a sort of tracking. This change will improve regular students' KS3 achievements by about 0.04 of a standard deviation (0.03+0.01), or 0.13 of a standard deviation (0.09+0.04) if we consider the within-pupil dispersion of KS3 achievements. Interestingly, this effect is consistent with the findings by Duflo et al. (2008) who document a 0.14 standard deviation improvement in the test score of pupils in primary schools in Kenya after 18 months of random assignment to homogenous (tracked) classes.

6.2. Effects of Peer Ability in Extended Samples

We noted earlier in the paper that the our 'natural experiment' sample is not representative of the whole population of students in England. To shed light on the external validity of our results, in Figures 3 and 4 we present estimates of the treatment effects alongside with confidence intervals (with standard errors clustered at the school level) obtained from 18 separate regressions that progressively use sub-samples with larger standard deviations of KS2 attainments across subjects. Note that these sub-sets of pupils are identical to those used to check balancing properties of our peer quality treatments (Figures 1 and 2), and move go from a s.d.$\leq$3 (including 6% of the sample) to a s.d.$\leq$11.5 (including around 50% of the full sample) in steps of 0.05. All regressions include pupil fixed effects, control for old peer quality and KS2 pupil own achievement, although our results are not sensitive to the inclusion of any of these controls.

The remarkable result emerging from Figure 3 is that the estimates obtained from various sub-samples, and in particular the two samples with s.d.≤3 and s.d.≤11.5, are not remarkably apart from each other, being respectively at 0.045 and 0.060. Additionally, their confidence intervals are largely overlapping, allowing us to reject the hypothesis that they are different. This evidence suggest that the imbalance in KS2 that emerges as we move to larger, less selected and more representative sample (see Figure 1), is too small to confound our estimates the effect of average peer quality on students' own KS3 test scores in a within-pupil regression.

Figure 4 presents similar evidence for the effects of 'good' peers (top 5%) and 'bad' peers (bottom 5%) in the 18 sub-samples described above. Again, we obtain a result that shows that the peer effects estimated from the sample with s.d.≤3 sample is virtually identical to the peer effects estimated from the sample with s.d.≤11.5 sample. Indeed, the top 5% estimated peer effects in these two sets of pupils are 0.068 and 0.065, respectively and the line connecting the point estimates of all the other sub-samples is almost horizontal throughout (and all estimates are significantly different from zero). Similarly, the bottom 5% estimated peer effects in the two samples are -0.145 and -0.121, for the smallest and largest sample respectively, and all the other estimates fall between these two. Once again, these findings are very important as they lend support to the robustness of our identification strategy. Moreover, they suggest that the estimates obtained using the 'natural experiment' sample have external validity, as they remain unaffected when we focus on the sample with s.d.≤11.5 sample, which includes about 50% of the whole population of English secondary school students and is representative of their average characteristics (see Appendix Table 2).

Note that we have carried out some additional exercises to check the external validity of our estimates based on the sample with s.d.≤3. In particular, rather than using cumulative bands of the standard deviation of KS2 attainments to select different sub-samples, we have used the following mutually exclusive bins: s.d.≤3 (6% of the sample); s.d.>3 and s.d.≤5.5 (approximately 10% of the sample); s.d.>5.5 and s.d.≤8 (approximately 13% of the sample); s.d.>8 and s.d.≤10.5 (approximately 13% of the sample), and finally s.d.>10.5 and s.d.≤11.5 (approximately 5% of the sample). Even in this case, we were unable to reject the null hypothesis that the estimated effects of the 'good', the 'bad' and the average peers are identical across the sub-samples.[9]

6.3. The Good, the Bad and the Average: Who has an Effect on Regular Students?

[9] The results are not displayed, but are available from the authors.

As set out in the Introduction, in our research we are interested in trying to disentangle which segments of the distribution of peer ability drive the impact of the average peer quality. Stated differently, we want to know the 'nature' of the ability peer effects estimated above, and understand where these might be coming from: do they originate from all parts of the peer ability distribution; or are the estimated effects mainly driven by the very weak and the very best students at school? To provide insight about this question, we estimate different models that includes all three measures of peer ability that we have used so far, namely, the 'good', the 'bad' and the average peer quality. Table 5 reports the regression coefficients and standard errors from a variety of specifications. All regressions include pupil fixed effects. Importantly, average peer quality in Panel A is calculated using all pupils in the sample, including those in the top and bottom 5% of the KS2 national distribution (as done so far). On the other hand, in Panel B, the average peer ability is calculated using only pupils in the full sample that are not in the top 5% nor in the bottom 5% of the KS2 distribution in any subject.

For both Panel A and Panel B, Columns 1 to 3 present estimates based on one single regression which simultaneously includes all peer quality measures. The results in the two panels are virtually identical, suggesting that it does not matter whether the average includes or excludes the top and bottom 5% new peers. Further, the estimates suggest conclusively that it is the top and bottom peers that have an impact on regular students in schools, and not the average peer quality. Indeed, the estimates of the effect of the top and bottom peers are very similar to those obtained when the average peer quality was not included in the specification (Table 4). On the other hand, the effect of the average peer quality is all but washed out in comparison to the our previous estimate when it was included as the only proxy of peers' ability (see Table 4).

In Columns 4 to 7 we report additional estimates where the average peer quality at KS2 is divided into two components: the average for pupils with KS2 percentile below the median (estimates reported in Column 4); and the average for pupils with KS2 percentile above the median (Column 5). The estimates in Columns 4 to 7 are therefore based on a regression that simultaneously includes four treatment measures of peer quality. These regression do not include, however, a control for previous peer quality in order to minimize the loss of observations, although the results when these are included are qualitatively and quantitatively similar.[10] The estimates presented in these four columns lend further support to the result that all the peer ability effects come from the tails of its distribution. The effects of the average quality of peers below and above the median are smaller the effects presented in Table 4, and not statistically significant at any conventional level, while the effects of the bottom 5% peers are only

[10] The loss in observation is due to the fact that for several pupils there are no old peers whose KS2 achievements are below or above the median of the KS2 distribution. Whenever this is the case, the average quality of old peers above/below the median of the KS2 distribution cannot be properly computed and it is assigned a missing value.

marginally reduced relative to the estimates in Columns 2 of the table, or in comparison to Table 4, and still highly significant. On the other hand, the estimates of the top 5% peers are smaller than before and only marginally significant, at approximately the 10% level. This is however due to the fact that this specification is rather demanding relative to the small size of our 'experimental' sample. Indeed, the estimates of the effect of the top 5% new peers obtained from this same specification on the sample with s.d.≤11.5 are strongly significant and of similar magnitude. All in all, the broad conclusion from this section is that peer effects at school mainly stem from very special students, namely those with very low ability and those with very high ability.

6.4. Falsification tests

Before presenting some results regarding the heterogeneity of our estimates, we perform two falsification tests to further assess the validity of our strategy. These boil down to 'mixing' our treatments across subjects in two different 'artificial' ways: in Falsification I, we move subjects so that Mathematics becomes English, Science becomes Mathematics, and English becomes Science; in Falsification II, Mathematics becomes Science, Science becomes English; and English becomes Mathematics. The results of these two falsification exercises are presented in Panel B of Table 4. Firstly, note that the OLS estimates are of the same sign, although of smaller magnitude, than the OLS estimates based on the true treatment measures. However, the within-pupil estimates are much smaller and have opposite sign than the true within-pupil estimates. For example, in Falsification I the within-pupil estimate of the average peer ability is -0.007 (se=0.014), while the estimate of the true treatment measure is 0.045 (se=0.016). The respective estimates for the effect of the bottom 5% peers are 0.029 (se=0.036) and -0.145 (se=0.040), while these estimates for the effect of the top 5% peers are -0.040 (se=0.031) and 0.068 (se=0.033). As seen from the comparison of these coefficients, the estimated standard errors are very similar (pair-wise), but the point estimates are very different (pair-wise). In Falsification II, the results are qualitatively similar although the erroneous treatments have more precise estimates than in Falsification I, giving rise to significant estimates but with signs that are, once again, opposite to those of the true treatment measures.

All in all, we interpret these falsification tests as compelling evidence that our main results do not capture a spurious correlation between the within-pupil variation in performance across subjects and the peer variation in performance across subjects. Overall, this evidence supports our interpretation of our findings in Table 4 and Table 5 as causal.

7. Allowing for Heterogeneous Effects

7.1. Heterogeneity by Students' Ability

In this part of our paper, we test for the presence of heterogeneous effects along a variety of dimensions. To start with, we look at whether very good and very bad peers differentially affect students with different academic abilities. For this purpose, we stratify the sample into five groups according to the distribution of the *average* of KS2 percentiles across subjects. The percentiles ranges defining the six non-overlapping groups are as follows: 5-20; 20-35; 35-50; 50-65; 65-80; and 80-95. Our regression models now include interaction terms of the percentages of top 5% peer and bottom 5% peer (separately for old and new peers) with a dummy indicating to which of the six non-overlapping KS2 ability groups a pupil belongs too. Note that main effects are included and that the effect of KS2 achievements is controlled for semi-parametrically by interacting pupil own KS2 percentiles with the dummies indicating his/her rank in the ability distribution.

Our estimates of the effects of the percentage of top 5% peer and the percentage of bottom 5% peer in interaction with treated pupils' ability groups are reported in the two Columns of Table 6. As before, these estimates are obtained from one single regression that includes both treatments.[11] Column 1 shows no clear variation in the negative effect of the bottom 5% across ability groups of regular students. Although all but one of the interaction terms are negative, none of them is significantly different from zero. It seems that the very weak students equally impact all other pupils. On the other hand, the interactions between the ability sub-groups of regular students and the top 5% new peers treatment show a different and interesting pattern. The positive effect of this treatment is seen in all groups, but for the very two top ability sets of regular students. In fact, for students in the group at the 65-80 percentiles the top 5% peer effect is practically zero (0.125-0.121), while for the students in the 80-95 percentiles group the effect is negative (0.125-0.247). This negative estimate implies that good pupils are negatively affected by having more good peers around them. Since it is a somewhat unusual finding we checked the robustness of this result and found that it persists even in the sample of s.d.<=11.5

What could explain this result? One possible explanation is that it is a purely mechanical effect: if we shift the ability distribution so as to have more of the very best top 5% students at school, this might crowd-out students who are in the next ability group (80-95 percentiles) from advanced courses or activities, such as Science and Mathematics clubs or special field trips because of limited space available in such activities.[12] In fact, using our data we can check whether there is any evidence related to this crowding-out mechanism. Assume that there is one top-tier class (tracking) for each subject in each

[11] Estimates of the average peer effect in the interaction with student ability are not reported, but are available upon requests. Broadly speaking, these effects did not show a clear pattern. The estimates for the percentiles 35-50 and 65-80 are positive and marginally significant, while estimates in the other ability ranges are mostly positive but not significant. As we already know that the average peer effects only reflect the impact of the very low and very high ability peers, we focus on these two.

[12] Along similar lines, an often used example would advocate that, if we shift the distribution of students from white to black, the probability of making the school basketball team goes down for both black and white students.

school. In this setting, having many good peers *in that subject*, would reduce one student's chances of getting into the top class, depress his/her achievements, and hence cause the finding of a negative effect of the top 5% of peers on the next most able students.[13] If this was the case, then our results should be sensitive to the overall cohort size: arguably, in smaller schools that have fewer classes per subject, crowding-out of the best class is going to be more costly since the quality-differentials of different classes should be larger.

To check for this possibility, we divide schools into four groups with different cohort-size and re-estimate our models. The four groups are the quartiles of the cohort size distribution and are defined as follows: below 163 pupils, 164 to 201 pupils, 201 to 237 pupils, and cohort-size above 237. Generally, we find that peer effects from the top 5% and bottom 5% pupils are estimated to be larger in smaller schools. This is not surprising given that pupils' interactions might be more frequent in schools with smaller cohorts. When we limit our sample to 'natural experiment' group of pupils with s.d.≤ 3, we find that the negative effect of the top 5% peers on pupils in the 80^{th}-95^{th} percentile of the ability distribution is only significant in the first two groups of smallest schools. To further probe our results in this direction, we then split schools into 10 cohort size deciles and extend the sample to s.d.<=11.5 in order to get larger samples and more precision within ability bands. When we do this, we find that the negative effect of top peers on the KS3 outcomes of pupils in the 80^{th}-95^{th} percentile of the ability distribution is not sensitive to the cohort size and – if anything – the effect seems somewhat weaker in smaller schools. Finally, we estimated a variety of models in samples stratified by school types, in particular by religious affiliation, and cohort size. Even then, we consistently found that the effect of the top 5% peers on the higher ability students remains negative. All in all, this evidence leads us to conclude that the crowding-out channel is not important here, although other possible explanations for this negative effect, such as students' competitive pressure, de-motivation and "big-fish-small-pond" mechanisms (Marsh, 2005), must remain conjectures given the information in our data.

Another possible and rather mechanical explanation for why pupils who are good on average suffer from having many top 5% peers might be related to mean-reversion. In general, the test score data do reveal significant mean reversion. For example, pupils in the 5^{th}-20^{th} percentile at KS2 experience a 4 percentile point average improvement in their KS3 test score. At the other end of the ability distribution, pupils in the 80^{th}-95^{th} KS2 percentile have an average 5.6 percentile deterioration in their average KS3. However, the within-pupil standard deviations in KS2 are (by construction) almost constant and constrained by ability bands, which means that pupils within the same ability group, in particular those in

13 Note that this result is at odds with the findings reported in Duflo et.al. (2008) that tracking in Kenyan primary schools is good for everyone and even for the 'marginal student'. However, this difference could be due to difference in context as the Kenya evidence is based on a very special type of primary schools, with only one multi age class (corresponding to 1st to 5th grade) in the school.

the 80^{th}-95^{th} KS2 percentile, would be similarly affected by mean-reversion, irrespective of how many good peers the interact with. Indeed, our identification relies on across subject differences in KS3 test scores and exposure to 'good' and 'bad' peers for the same pupil. Hence it is not clear why mean reversion should be causing the peculiar pattern of heterogeneous effect by student's ability documented in Table 6. Moreover, if mean reversion was to explain our findings, there is no reason to believe this should only affect the top of the ability distribution, and not the bottom as well. However, we do not observe any significant interaction between either the top 5% peers or the bottom 5% peers and the fact that a student ranks low in the KS2 ability distribution (e.g., in the 5^{th}-15^{th} KS2 percentile).

Nevertheless, to shed further light on this issue, we check whether the pure effect of belonging to – say – the top-group in the average KS2 ability distribution (80^{th}-95^{th} percentile) is related to the KS3 outcomes of students. Our results reveal that there is evidence that simply belonging to the top 15% of KS2 ability distribution of regular students negatively affect KS3 outcomes, nor do we find any relation for any other ability group. In a nutshell, it is only the interaction between belonging to the top ability group and having many new 'good' peers that gives rise to the negative coefficients discussed here above. Our conclusion therefore is that, although there is significant mean reversion in average scores, this cannot be driving our findings.

7.2. Gender Heterogeneity in Treatment Effects

Naturally, the heterogeneity of peer effects by gender is also interesting, especially in secondary school where the social interactions between boys and girls intensified. In Table 7, we report some results based on separate samples for boys and girls and where the peer quality measures capture 'treatment' from the very weak pupils (bottom 5%) and the very best ones (top 5%). The first row reports the estimates of the overall effect in each sample by gender. We find that the effect of the bottom 5% peers is negative in both gender groups, although is it much smaller for boys (at -0.065) and hardly significant. Additionally, the effect of the top peer group on girls is positive, significant and sizeable at 0.077 (this is comparable to the effect reported in Table 4. On the other hand, the effect of the top 5% new peers on boys is negative, at -0.079, and marginally significant.

To further investigate this issue, we study the sign and size of ability peer effects separately of boys and girls *and* in interaction with students' own ability (as in Table 6). Our results are tabulated in Panel B of Table 7. Firstly, we note that our finding of a negative effect of top peers on boys is evident for almost all ability groups, but it is largest for the most able among the regular students (and significant for this group only). For girls, the effect of top peers is positive for almost all ability groups, but it is significantly negative for the most able students. All in all, it seems like a larger representation of top 5% peers is less beneficial or even harmful for the best regular students, irrespective of their gender.

Since these results are somewhat unexpected, in particular the finding that boys are negatively affected by having a high proportion of very bright students at school, we performed a series of checks to assess to robustness of our findings. Firstly, we estimated our models using bigger samples, in particular by including boys and girls with s.d.<=11.5. The same patterns emerged when the estimates were obtained from this larger and more representative set of pupils: the negative effect of the top 5% peers on boys is present for most of the ability distribution of regular students, whereas for girls this is still only present at the very top. We also pondered whether one possible explanation for this result is that there are too few boys relative to girls at the top of the ability distribution to properly estimate separate effects for boys and girls in different ability groups, but this does not seem to be the case. Another possible explanation is that the negative effect is 'mechanical', and once more due to mean-reversion or a ceiling effect. However, we have already ruled our this explanation in general for the findings discussed in Section 7.1 and there is no valid reason why our arguments would apply differently for boys and girls.

Therefore, a natural conclusion is that these effects might be 'real' and the question is whether this pattern of heterogeneity by gender can be related to other findings in the literature. The stronger effects on girls that we find here are consistent with growing evidence that girls are more affected than boys by education inputs and interventions. For example, Anderson (2008) shows that three well-known early childhood interventions (namely, Abecedarian, Perry and the Early Training Project) had substantial short and long term effects on girls, but no effect on boys. Likewise, the Moving to Opportunity randomized evaluation of housing vouchers generated clear benefits for girls, with little or even adverse effects on boys (Katz et al., 2001). Moreover, some recent studies show a consistent pattern of stronger female response to financial incentives in education, with the evidence coming from a surprising variety of settings. For example, Angrist and Lavy (forthcoming), report larger effects of achievement incentives for girls in high schools in Israel than for boys. Closely related is a recent randomized trial looking at cash payments for academic achievement among college freshman: this study also finds clear effects for females, but no effect on males (Angrist et al., 2009).[14] Finally, a number of public-sector training programs generated larger effects on women than men (Lalonde, 1995).

7.3. Heterogeneity in Treatment Effects and Measures by Family Economic Status

We conclude our paper by studying whether we can detect some heterogeneity in ability peer effects by pupils' family economic status which we measure by eligibility for free school meals (FSM), a proxy for family income. In particular, we are interested in two questions: whether our ability peer effects are different for pupils with different family background, and by different levels of ability, and whether the

[14] Dynarski and Scott-Clayton (2008) on tuition aid and Garibaldi et al. (2006) on tuition penalties also find larger effects for girls than for boys.

eligibility for free school meals (or not) of the top 5% and bottom 5% peers themselves gives rise to some heterogeneous effects. The answer to the latter question can help us disentangle whether our main findings are driven by the academic ability of peers or by their family background.

In Table 8, we present some evidence to answer the first question. The structure of this table is identical to that of Table 7, except that we now split students into those eligible for FSM and those who are not. Panel A presents some estimates obtained by pooling pupils of all ability groups. The results for pupils non-eligible for FSM are very similar to those obtained in Table 4 over all students. More interestingly, we find that pupils eligible for FSM are affected by a greater margin by both the top 5% and bottom 5% peers, relative to our previous findings in Table 4. Note that this is not just because the within-pupil standard deviation in the two treatments is different for this sub-set of pupils relative to the rest of the population.

To shed more light on these findings, in Panel B of the table we break down our treatment effect estimates by pupils' own ability, and again separately according to pupils' eligibility for FSM. The results show that our top 5% and bottom 5% peer effects estimates are roughly constant throughout the ability distribution of regular students who are eligible for FSM. Importantly, the negative effect of the very bright peers on the most able regular pupils documented in Tables 6 and 7 completely disappears among pupils from more disadvantaged family background, and this is not just because of lack of precision in our estimates: the sign of the interaction effect between being exposed to many talented students and being in the 80^{th}-95^{th} percentile of the KS2 test score distribution is positive, although not significant. On the other hand, we still find that a higher fraction of top 5% peers has a negative impact on the KS3 attainments of pupils who are not eligible for FSM and rank in the highest percentiles of the ability distribution, but the overall effect remains positive and significant in the bottom 50 percent of the ability distribution. Finally, the negative effect of being exposed to many bed peers is broadly negative across the ability distribution for pupils non-eligible for FSM.

In Table 9, we present some findings that are related to whether eligibility for FSM (or not) of the top 5% and the bottom 5% peers affects the estimates of the treatment effects. The table has three panels: Panel A considers all pupils; Panel B focuses on pupils eligible for FSM; and Panel C looks at pupils non-eligible for FSM. More importantly, our treatments are now split into to four variables: the fraction of bottom 5% peers eligible for FSM (Column 1); the fraction of bottom 5% peers non-eligible for FSM (Column 2); the fraction of top 5% peers eligible for FSM (Colum 3); and the fraction of top 5% peers non-eligible for FSM (Column 4). Note that the treatment variables constructed separately for FSM-eligible pupils and non-FSM eligible pupils are based on relatively few observations, in particular our proxy for the very good FSM-eligible peers and the very bad non-FSM-eligible peers. This is because on average only 15% of pupils are eligible for FSM, and they tend to be under-represented in the top 5% of the KS2 distribution (only 5%; see Table 1) and over-represented in the bottom 5% of the KS2 distribution (up to 30%; see Table 1). Therefore we expect our results to be quite ‘noisy’ and not very precisely estimated. Note also that the coefficients tabulated in each panel are obtained from one single regressions containing all four treatment variables.

Starting from Panel A, we do find some evidence that ‘good’ peers exert a positive effect, while ‘bad’ peers exert a negative one, on other pupils’ achievements irrespective of their family background. There is clear evidence that peers in the bottom 5% of the ability distribution have a negative and significant impact on the KS3 achievement of regular students, no matter whether the peers are eligible for FSM or not. However, the treatment estimate of bottom 5% pupils from low income families is larger, -0.225 versus -0.072, though these impact needs to be re-scaled to account for the higher within-pupil standard deviation of the treatment variable in Column (1). Interestingly, moving down the panels of the table, we find that these results hold when we consider separately pupils who are eligible for FSM and those who are not. Although the effect of the bottom 5% peers who are not-eligible for FSM on pupils who themselves are not eligible for FSM is not precisely estimated, the sign and size of the coefficients point in the same direction, and we believe the lack of statistical significance is due to the fact that there are very few ‘bad’ peers who are not on FSM, in particular when we focus on schools that might contain a larger fraction of pupils from better family background (as implicitly done in Panel C). On the other hand, the evidence on the effect of the very talented peers on other students is more mixed. Throughout the three panels of the table, we find a positive and significant effect of the top 5% pupils who are not eligible for FSM. However, the estimated impact of the very bright students eligible for FSM on the KS3 attainment of regular pupils is not significant. Nevertheless, the size of the coefficient and its sign do suggest that very talented pupils still exert a positive effect on other students’ KS3 test scores even when they come from a disadvantaged background. Once again, we attribute the lack of precision to the fact that there are very few FSM-eligible pupils in the top 5% of the KS2 test score distribution, which implies that our proxies are very noisy. In conclusion, we believe the results in this section suggest that overall the positive effect of ‘good’ peers and the negative effect of ‘bad’ one are driven by peers’ academic ability, and not related to their family background.

8. Conclusions

In this paper, we have estimated ability peer effects in schools using data for all secondary schools in England for four cohorts of age-14 (9^{th} grade) pupils and measuring the peers’ quality by their academic ability as recorded by test scores at age-11 (6^{th} grade). Importantly, in order to shed some light on the nature of peer effects, we have estimated both the effect of the average ability of peers, as well as the effect of being at school with a high proportion of very low-ability students or very high-ability ones, on the cognitive outcomes of regular students.

We view our main methodological contribution as twofold. Firstly, we are able to measure peer ability by test scores that directly capture the cognitive ability of pupils and that are pre-determined with respect to peer interactions in secondary schools, since they are measured at the end of primary education before pupils change schools to start their secondary education. Moreover, by focusing only on peer quality measures that are based on new peers in secondary schools we were able to completely by-pass reflection problems. Secondly, we offer a new approach to measuring peer effects, by focusing on within-pupil variation in performance across multiple subjects in a setting where peers' quality is also measured by the variation in their ability across subjects. By using student fixed-effect estimation we are simultaneously able to fully control for family and school fixed unobservables, and pupil ability that is constant across subjects. As some degree of sorting and selection is still evident in our data even when we look at the correlation between within-pupil variation and within-peers variation in ability across subjects, we proposed to focus on a 'natural experiment' sample of pupils with little or no differences in prior ability across subjects as measured by their achievements at the end of primary school. In this 'experimental' sample there can be no relationship between within-pupil variation and within-peers variation in ability because, by construction, the former exhibit no variation. This limits to the minimum that chances that our estimates are biased because of endogenous sorting of pupils or omitted variables.

Our results show that higher peer average ability at school has a positive and significant effect on the achievements of other students. Additionally, we find that a high concentration of very low ability students ('bad' peers) significantly lowers the academic achievements of regular students, while a high concentration of very high ability students ('good' peers) significantly increases their academic achievements. More importantly, the effect of average peer quality is dominated by the effect of the very bright and the very worst peers, which suggests that what matters in the transmission of peer effects in secondary schools in England is the concentration of exceptionally able or weak schoolmates. We also identify and discuss the heterogeneity of peer effects along a variety of dimensions. One striking result is that the very brilliant pupils at school negatively impact the academic performance of boys, and in particular those who are among the second highest group at school in terms of ability. On the other hand, girls benefit more from having high achievers at school, although there is some evidence that the highest ability girls among regular students at school benefit the least (or even loose out) from these interactions. Additionally, we find that our results are very similar for pupils eligible for free school meals or not, and that the effects of the very bright and very weak schoolmates are not dramatically affected if we break down our measures of peers quality according to their free school meals eligibility. This backs our intuition that our peer quality measures mainly capture academic ability, and that our findings are not driven by schoolmates' family background.

As a general remark, the findings that we have presented in this paper are important because of their strong external validity. Our full data includes over 90 percent of four cohorts of pupils in England that transit from primary school through to the third year of secondary schooling, and sit for two crucial standardized national tests, namely the Key Stage 2 (6^{th} grade) and Key Stage 3 (9^{th} grade). While most of our estimates come from the 'natural experiment' group of pupils with little variation in test scores at the end of primary school, our findings are unaffected when we look at large and more representative sub-set of pupils. In particular, when we look at students with a standard deviation of Key Stage 2 test scores below 11 points, we confirm that our estimates remain virtually identical. This is very important as this sample accounts for 50 percent of the whole population of secondary state-school students in England and has characteristics that are representative of those of English student nationwide. Additionally, our various samples are large enough to allow us to recover a variety of estimates about the heterogeneity of our treatment effects. In this respect, our paper is a direct response to some of the concerns raised by Deaton (2009) about the limitations of randomized controlled trials in terms of recovering heterogeneous effects and having strong external validity.

Finally, the peer effect estimates that we present in this paper are of reasonable size. In comparison to the effects of classroom instructional time estimated in Lavy (2009), changing the peer environment from the worse to the best observed in English secondary schools would be equivalent to two more weekly hours of instructional time. Our estimates also imply that if schools were organized in a way to include all pupils, but the very bright and the very weak (i.e. a sort of tracking), the change in achievements would be equivalent to the effects of full tracking of pupils by ability based on the experimental evidence presented in Duflo et a. (2008), with the caveat that their findings come from multi-age classes in primary schools in Kenya.

Do our results overall lend support to tracking of students by ability? Besides any equity consideration, there is no simple answer to this question from an efficiency point of view. As already mentioned, making schools more homogeneous by excluding both very good and very bad peers would result in an overall improvement in students' performance. However, the results are quite heterogeneous according to one pupils' own ability, gender and eligibility for free school meals. For example, pupils at the bottom of the ability distribution would not necessarily benefit nor loose out from more homogenous schools, as the beneficial effects of the top 5% peers almost perfectly counterbalance the negative peer effect of the bottom 5% students. The opposite is true pupils at the top of the ability distribution, in particular for girls, who would significantly *gain* from not interacting with the very weak and very academically bright students. Similarly heterogeneous conclusions can be reached when looking at pupils eligible for free-school meals or not. In conclusion, we believe our findings, despite not providing a one-size-fit-all policy recommendation, are rich enough to provide a solid ground for insightful interventions targeting students' ability mix as a means to improve learning standards.

References

Aizer A., (2008): "Peer Effects and Human Capital Accumulation: The Externalities of ADD," Brown University, mimeo.

Ammermueller, A and J. Pischke (2006): "Peer Effects in European Primary Schools: Evidence from PIRLS," IZA Discussion Paper 2077.

Anderson, M. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103(484), 1481-1495.

Angrist, J.D. and K. Lang, (2004): "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review*, 94(5), 1613-1634.

Angrist, J.D., D. Lang and P. Oreopolous (2009): "Incentives and Services for College Achievement: Evidence from a Randomized Trial", *American Economic Journal: Applied Economics*, 1.

Angrist, J.D. and V. Lavy (forthcoming): "The Effect of High-Stakes High School Achievement Awards: Evidence from a Group-Randomized Trial," *American Economic Review*.

Arcidiacono, P. and S. Nicholson (2005): "Peer Effects in Medical School," *Journal of Public Economics*, 89, 327-350.

Bayer, P., R. Pintoff and D. Pozen, (2004): "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections," Yale University, mimeo.

Becker, G.S. (1974): "A Theory of Social Interactions," *Journal of Political Economy*, vol. 82(6), 1063-93.

Bifulco, R, J. Fletcher and S. Ross (2008): "The Effect of Classmate Characteristics on Individual Outcomes: Evidence from the Add Health," University of Connecticut, Department of Economics Working Paper 2008-21.

Black, S.E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics,* 114(2), 577–99.

Burke, M.A. and T.R. Sass (2004): "Classroom Peer Effects and Student Achievement," Florida State University, mimeo.

Carrell, S. E. and M.L. Hoekstra (2008): "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," paper presented at the NBER Summer Institute, Labor Studies.

Coleman, J. (1966): "Equality of Educational Opportunity," Washington D.C., U.S., Department of Health, Education and Welfare, Office of Education, OE-38001.

Deaton, A. (2009): “Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development”, Research Program in Development Studies, Center for Health and Wellbeing, Princeton University, mimeo.

Duflo E., P. Dupas and M. Kremer (2008): “Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya”, MIT, mimeo.

Dynarski, S. and J. Scott-Clayton (2008): “Complexity and Targeting in Federal Student Aid: A Quantitative Analysis,” *Tax Policy and the Economy*, 22, J. Poterba (ed), Cambridge, MA: NBER.

Figlio, D.N. (2007): “Boys Named Sue: Disruptive Children and Their Peers,” *Education Finance and Policy*, 2(4), 376-394.

Garibaldi, P., F. Giavazzi, A. Ichino and E. Rettore (2006): “College Cost and Time to Complete a Degree: Evidence from Discontinuities”, University of Bologna, mimeo.

Gibbons, S., S. Machin and O. Silva (2009): “Valuing School Quality Using Boundary Discontinuity Regressions”, London School of Economics, mimeo.

Gibbons, S and S. Telhaj (2008): “Peers and Achievement in England's Secondary Schools,” SERC Discussion Paper 1.

Gould, E.D., V. Lavy and D.M. Paserman (forthcoming): “Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence,” *Economic Journal*.

Goux, D. and E. Maurin (2007): “Close Neighbours Matter: Neighbourhood Effects on Early Performance at School,” *Economic Journal*, 117, 1-24.

Hanushek, E., J. Kain, J. Markman and S. Rivkin (2003): “Does Peer Ability Affect Student Achievement?,” *Journal of Applied Econometrics*, 18(5), 527-544.

Hoxby, C.M. (2000): “Peer Effects in the Classroom: Learning from Gender and Race Variation,” NBER Working Paper 7867.

Kane, T., J. Riegg, S.K. and D.O. Staiger (2006): “School Quality, Neighborhoods and Housing Prices,” *American Law and Economics Review*, 8(2), 183-212.

Katz, L.F., J.R. Kling and J.B. Liebman (2001): “Moving to Opportunity in Boston: Early Results from a Randomized Mobility Experiment,” *Quarterly Journal of Economics*, 116, 607-654.

LaLonde, R.J. (1995): “The Promise of U.S. Employment and Training Programs,” *Journal of Economic Perspectives*, 9(2), 149-168.

Lavy, V. (2009): “The Causal Effect of Instructional Time on Achievements in Math, Science and Reading: International Evidence”, paper presented at the Economics of Education and Education Policy in Europe (EEEPE) Network Final Conference.

Lavy, V. and A. Schlosser (2008): “Mechanisms and Impacts of Gender Peer Effects at School,” NBER Working Paper 13292.

Lavy, V., D.M. Paserman and A. Schlosser (2008): “Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom,” NBER Working Paper 14415.

Lefgren, L. (2004): “Educational Peer Effects and the Chicago Public Schools,” *Journal of Urban Economics*, 56(2), 169-191.

Manski, C. (1993): “Identification of Endogenous Social Effects: The Reflection Problem,” *Review of Economic Studies*, 60(3), 531-542.

Marsh, H. (2005): “Big Fish Little Pond Effect on Academic Self-concept: Cross-cultural and Cross-Disciplinary Generalizability”, paper presented at the AARE Conference Paper.

Moffit, R. (2001): “Policy Interventions, Low-Level Equilibria, and Social Interactions”, in *Social Dynamics*, S. N. Durlauf and H. Young (eds.), Cambridge MA: MIT, 45-82.

Nechyba, T. and J. Vigdor (2005): “Peer Effects in North Carolina Public Schools.” Duke University, mimeo.

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

Sanbonmatsu, L., J.R. Kling, G.J. Duncan and J. Brooks-Gunn (2004): “Neighbourhoods and Academic Achievement: Results from the Moving to Opportunity Experiment,” Industrial Relation Section, Princeton University Working Paper, 492.

Vigdor, J. and T. Nechyba (2004): “Peer Effects in Elementary School: Learning from ‘Apparent’ Random Assignment,” Duke University, mimeo.

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

Tables

Table 1 – Descriptive statistics: pupils' outcomes, pupils' background and school characteristics

Variable	Full sample	Selected sample (std.dev.≤3)	At least 1 subject top 5%	At least 1 subject bottom 5%
Panel A: Pupils' outcomes				
KS2 percentile, English	49.3 (24.3)	42.0 (29.8)	87.1 (14.8)	8.5 (12.5)
KS2 percentile, Mathematics	49.4 (24.3)	42.0 (29.9)	87.0 (14.1)	9.4 (13.6)
KS2 percentile, Science	48.9 (24.3)	41.9 (29.9)	87.7 (13.1)	10.9 (15.5)
KS3 percentile, English	48.9 (26.0)	43.0 (29.4)	81.2 (18.6)	15.3 (18.2)
KS3 percentile, Mathematics	49.2 (25.3)	42.6 (29.8)	84.5 (16.3)	14.8 (17.6)
KS3 percentile, Science	49.2 (25.5)	43.2 (29.9)	84.4 (16.2)	16.0 (17.9)
Panel B: Pupils' characteristics				
First language is English	0.93 (0.253)	0.93 (0.26)	0.95 (0.21)	0.89 (0.31)
Eligible for free school meals	0.13 (0.337)	0.16 (0.36)	0.05 (0.22)	0.30 (0.46)
Male	0.50 (0.500)	0.49 (0.50)	0.48 (0.50)	0.55 (0.50)
Changed school between Year 7 and KS3	0.11 (0.313)	0.11 (0.32)	0.09 (0.29)	0.14 (0.35)
Ethnicity: White British	0.85 (0.35)	0.85 (0.36)	0.88 (0.32)	0.81 (0.39)
Ethnicity: White other	0.02 (0.12)	0.02 (0.13)	0.02 (0.13)	0.02 (0.14)
Ethnicity: Asian	0.05 (0.22)	0.05 (0.22)	0.034 (0.18)	0.07 (0.26)
Ethnicity: Black	0.03 (0.16)	0.03 (0.16)	0.01 (0.11)	0.04 (0.19)
Ethnicity: Chinese	0.00 (0.05)	0.00 (0.04)	0.00 (0.07)	0.00 (0.04)
Ethnicity: Other	0.05 (0.22)	0.05 (0.22)	0.07 (0.21)	0.06 (0.24)
Panel C: School characteristics (Year 7)				
Cohort size	201.7 (57.2)	201.9 (57.4)	204.1 (56.3)	198.8 (58.5)
Community school	0.67 (0.47)	0.68 (0.47)	0.63 (0.48)	0.73 (0.44)
Voluntary aided school	0.14 (0.35)	0.14 (0.34)	0.17 (0.38)	0.10 (0.30)
Voluntary controlled school	0.03 (0.18)	0.03 (0.18)	0.04 (0.19)	0.03 (0.17)
Foundation school	0.15 (0.36)	0.15 (0.36)	0.16 (0.36)	0.13 (0.34)
City Technology college school	0.00 (0.05)	0.00 (0.05)	0.00 (0.07)	0.00 (0.03)
Religiously affiliated school	0.16 (0.37)	0.15 (0.36)	0.19 (0.39)	0.11 (0.32)

Note: Table report means of the listed variables and standard deviation in parenthesis. Number of pupils in full sample: 1,279,514. Number of pupils in selected sample: 77,322. Selected sample is composed of pupils with standard deviation of KS2 percentiles across subjects ≤3. Full sample and selected sample only include pupils with KS2 achievement in each subject above the 5th percentile and below 95th percentile of KS2 cohort-specific national distribution. Number of pupils with at least one subject in top 5% (≥95th percentile of KS2 cohort-specific national distribution): 172,634. Number of pupils with at least one subject in bottom 5% (≤5th percentile of KS2 cohort-specific national distribution): 130,459. Year 7 refers to the first year in secondary school after transition out of primary. KS3 refers to Year 9 when pupil sit for their KS3 assessment. Fractions may not sum to 1. This is due to rounding or partially missing information.

Table 2 – Descriptive statistics of treatments: percentages of pupils in top and bottom 5% of KS2 ability distribution and average KS2 achievements (percentiles)

	Full sample				*Selected sample (std.dev.≤3)*			
Variable	Mean	Std. dev.	Min	Max	Mean	Std. dev.	Min	Max
Panel A: Average KS2 percentile treatments								
Average peer achievement at KS2 in English, *all peers*	49.99	7.55	22.70	75.34	49.51	7.64	23.94	75.23
Average peer achievement at KS2 in English, *new peers*	49.79	8.71	1	98	49.42	8.71	1	98
Average peer achievement at KS2 in Maths, *all peers*	50.16	6.85	23.94	72.82	49.74	6.95	24.77	72.64
Average peer achievement at KS2 in Math, *new peers*	49.94	8.06	1	100	49.62	8.08	1	100
Average peer achievement at KS2 in Science, *all peers*	49.93	7.35	25.31	73.83	49.46	7.42	25.54	73.50
Average peer achievement at KS2 in Science, *new peers*	49.68	8.35	1	100	49.34	8.37	1	100
Panel B: Top 5% treatments								
Fraction of top 5% in English, *all peers*	5.14	3.19	0	19.56	5.01	3.17	0	19.56
Fraction of top 5% in English, *new peers*	4.22	3.03	0	19.56	4.13	3.01	0	19.55
Fraction of top 5% in Maths, *all peers*	4.63	2.70	0	19.87	4.55	2.69	0	19.87
Fraction of top 5% in Maths, *new peers*	3.77	2.60	0	19.87	3.72	2.58	0	19.87
Fraction of top 5% in Science, *all peers*	4.84	2.92	0	19.86	4.72	2.89	0	19.86
Fraction of top 5% in Science, *new peers*	3.91	2.75	0	19.86	3.83	2.72	0	19.82
Panel C: Bottom 5% treatments								
Fraction of bottom 5% in English, *all peers*	4.64	3.00	0	19.30	4.80	3.07	0	19.30
Fraction of bottom 5% in English, *new peers*	3.79	2.78	0	19.30	3.90	2.84	0	19.30
Fraction of bottom 5% in Maths, *all peers*	4.68	2.86	0	19.86	4.84	2.93	0	19.86
Fraction of bottom 5% in Maths, *new peers*	3.81	2.67	0	19.86	3.91	2.72	0	19.48
Fraction of bottom 5% in Science, *all peers*	4.59	3.10	0	19.78	4.76	3.19	0	19.78
Fraction of bottom 5% in Science, *new peers*	3.78	2.90	0	19.78	3.89	2.95	0	19.23
Percentages of new peers for pupils in Year 7	87.56	22.66	0	1	87.50	22.55	0	1

Note: Treatment measured in Year 7 when students start secondary school after transition from primary. 'All peers' refer to all students in the cohort in Year 7. 'New peers' refers to students in Year 7 in a given cohort that do not come from the same primary school. Average KS2 percentiles of peers always computed excluding the pupil under analysis.

Table 3 – Balancing of individual characteristics with respect to treatments, selected sample (std.dev.≤3)

	Average achievement at KS2 (percentiles), new peers	*Percentage of top 5% pupils, new peers*	*Percentage of bottom 5% pupils, new peers*
Dependent variable is:	(1)	(3)	(5)
Panel A: OLS regression results			
First language is English	0.003	-0.014	0.001
	(0.000)**	(0.001)**	(0.001)
Eligible for free school	-0.004	0.020	-0.006
meals	(0.000)**	(0.001)**	(0.001)**
Male	0.000	-0.001	0.001
	(0.001)	(0.001)	(0.001)
Changed school between Year 7 and KS3	-0.004	-0.011	-0.017
	(0.000)**	(0.001)**	(0.001)**
Ethnicity: White British	0.003	-0.017	0.000
	(0.000)**	(0.002)**	(0.001)
Ethnicity: White other	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)
Ethnicity: Asian	-0.002	0.011	0.000
	(0.000)**	(0.001)**	(0.001)
Ethnicity: Black	-0.001	0.004	-0.001
	(0.000)**	(0.001)**	(0.001)
Ethnicity: Chinese	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.001)
Ethnicity: Other	-0.000	0.001	0.000
	(0.001)	(0.000)*	(0.001)
KS2 percentiles	0.291	-1.155	1.094
	(0.015)**	(0.044)**	(0.047)**
Panel B: Within-pupil regression results			
KS2 percentiles	0.002	-0.006	0.005
	(0.002)	(0.005)	(0.004)
Controlling for old peers quality	Yes	Yes	Yes

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the effects of the percentage of top 5% pupils and percentage of bottom 5% pupils are obtained from one single regression including both treatments. Estimates of the effect of average peer achievements at KS2 obtained from a separate regression. The table displays the coefficients on treatments based on new peers; all regressions control for quality of old peers. Standard error clustered at the school level as identified in Year 7. **: at least 1% significant; *: at least 10% significant. Number of observations in Column 1 is 218,184, corresponding to 72,728 pupils; number of schools: 2193. Number of observations in Columns 2 and 3 is 231,966, corresponding to 77,322 pupils; number of schools: 2194.

Table 4 – Impact of peer quality on KS3 educational attainments (percentiles), selected sample (std.dev.≤3)

	Average achievement at KS2 (percentiles), new peers		*Percentage of bottom 5% pupils, new peers*		*Percentage of top 5% pupils, new peers*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
Panel A: Main findings						
KS3 percentiles	0.444	0.046	-1.225	-0.148	1.088	0.071
	(0.018)**	(0.016)**	(0.045)**	(0.040)**	(0.047)**	(0.033)*
KS3 percentiles, controlling for KS2	0.224	0.045	-0.486	-0.145	0.324	0.068
	(0.010)**	(0.016)**	(0.026)**	(0.040)**	(0.026)**	(0.033)*
Panel B: Falsification tests						
KS3 percentiles, controlling for KS2:	0.207	-0.007	-0.441	0.029	0.314	-0.040
falsification exercise I	(0.010)**	(0.014)	(0.026)**	(0.036)	(0.025)**	(0.031)
KS3 percentiles, controlling for KS2:	0.202	-0.037	-0.437	0.115	0.304	-0.027
falsification exercise II	(0.010)**	(0.014)**	(0.027)**	(0.045)**	(0.025)**	(0.029)
Controlling for pupil characteristics	Yes	Absorbed	Yes	Absorbed	Yes	Absorbed

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the effects of the percentage of top 5% pupils and percentage of bottom 5% pupils are obtained from one single regression including both treatments. Estimates of the effect of average peer achievement at KS2 obtained from a separate regression. The table displays the coefficients on treatments based on new peers. Regressions in Row 1, 2, 4 and 5 control for quality of old peers; regressions in Row 3 does not control for old peer quality. Falsification exercise I mixes treatments across subjects as follows: Maths becomes English; Science becomes Maths; and English becomes Science. Falsification exercise II mixes treatments across subjects as follows: Maths becomes Science; Science becomes English; and English becomes Maths. Standard error clustered at the school level as identified in Year 7. **: at least 1% significant; *: at least 10% significant. Number of observations in Columns 1 and 2 is 218,184, corresponding to 72,728 pupils; number of schools: 2193. Number of observations in Columns 3 to 6 is 231,966, corresponding to 77,322 pupils; number of schools: 2194. Pupil characteristics include: First language is English; Eligible for free school meals; Male; Changed school between year 7 and KS3; Ethnicity: White other; Ethnicity: Asian; Ethnicity: Black; Ethnicity: Chinese; Ethnicity: Other.

Table 5 – Impact of peer quality on KS3 educational attainments (percentiles): top, bottom and average peer quality simultaneously, selected sample (std.dev.≤3)

	Average peer KS2, new peers	*Bottom 5% pupils, new peers*	*Top 5% pupils, new peers*	*Average peer KS2, new peers: below median*	*Average peer KS2, new peers: above median*	*Bottom 5% pupils, new peers*	*Top 5% pupils, new peers*
Dependent variable is:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Average peer quality includes top/bottom 5% pupils							
KS3 percentiles, controlling for KS2	0.019	-0.147	0.070	0.025	0.011	-0.131	0.058
	(0.016)	(0.041)**	(0.035)*	(0.021)	(0.019)	(0.041)**	(0.035)
Panel B: Average peer quality excludes top/bottom 5% pupils							
KS3 percentiles, controlling for KS2	0.025	-0.148	0.069	0.019	0.026	-0.137	0.055
	(0.014)	(0.041)**	(0.035)*	(0.022)	(0.021)	(0.040)**	(0.034)

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. All regression include pupils' fixed effects. Average peer quality in Panel A is calculated using all pupils in the sample, including those in the top 5% and bottom 5% of the KS2 cohort-specific national distribution. Average peer quality in Panel B is calculated using pupils in the full sample that are not in the top 5% or in the bottom 5% of KS2 cohort-specific national distribution in any subject. Average peer quality at KS2 broken down in Columns 4 and 5 into: average for pupils with KS2 percentile below median; average for pupils with KS2 percentile above the median. Estimates for Columns 1 to 3 coming from one single regression including simultaneously all peer quality measures; estimates for Columns 4 to 7 coming from another regression including simultaneously all peer quality measures. Regressions for results in Columns 1 to 3 include control for old peer quality; regressions for results in Columns 4 to 7 do not include old peer quality. This is to minimize the loss of observations (see body text for a discussion). Results including old peer quality measures in Columns (4) to (7) are qualitatively and quantitatively similar. Standard error clustered at the school level as identified in Year 7. **: at least 1% significant; *: at least 10% significant. Number of observations varies depending on specification (approx. 228,000 observations over 76,000 students in 2170 schools).

Table 6 – Impact of peer quality on KS3 attainments (percentiles), by ability, selected sample (std.dev.≤3)

	Percentage of bottom 5% pupils, new peers	*Percentage of top 5% pupils, new peers*
Dependent variable is:	(1)	(2)
KS3 percentiles, controlling for KS2 (baseline)	-0.139	0.125
	(0.040)**	(0.038)**
× *Percentile 20-35*	-0.001	0.100
	(0.085)	(0.076)
× *Percentile 35-50*	-0.005	0.026
	(0.012)	(0.093)
× *Percentile 50-65*	-0.037	-0.057
	(0.130)	(0.095)
× *Percentile 65-80*	-0.113	-0.121
	(0.139)	(0.089)
× *Percentile 80-95*	0.068	-0.247
	(0.098)	(0.066)**
Controlling for old peers quality (and interactions)	Yes	Yes

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the effects of the percentage of top 5% pupils and percentage of bottom 5% pupils are obtained from one single regression including both treatments. The table displays the coefficient on treatments based on new peers. All regressions control for the quality of old peers. Number of observations in Column 1 and 2 is 231,966, corresponding to 77,322 pupils; number of schools: 2194. Standard error clustered at the school level as identified in Year 7. **: at least 1% significant; *: at least 10% significant. Interaction terms obtained by interacting the percentage of top 5% pupils and percentage of bottom 5% pupils (separately for old and new peers) with a dummy indicating where the pupil ranks in terms of his/her KS2 percentiles *on average across subjects*. The effect of KS2 achievement is controlled for semi-parametrically by interacting pupil KS2 percentiles with the dummies indicating his/her rank in the ability distribution. Main effects included. Ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution.

Table 7 – Impact of peer quality on KS3 attainments (percentiles), by ability and gender, selected sample (std.dev.≤3)

	Boys only		Girls only	
	Percentage of bottom 5% pupils, new peers	*Percentage of top 5% pupils, new peers*	*Percentage of bottom 5% pupils, new peers*	*Percentage of top 5% pupils, new peers*
Dependent variable is:	(1)	(2)	(3)	(4)
Panel A: Pupils of ability pooled (overall effect)				
KS3 percentiles, controlling for KS2 (baseline)	-0.065	-0.079	-0.117	0.077
	(0.049)	(0.040)*	(0.049)*	(0.040)*
Panel B: Ability blocks defined on original KS2 percentiles				
KS3 percentiles, controlling for KS2 (baseline)	-0.096	0.037	-0.076	0.141
	(0.051)*	(0.049)	(0.053)	(0.049)**
× *Percentile 20-35*	0.046	-0.177	-0.086	0.313
	(0.119)	(0.098)	(0.115)	(0.103)**
× *Percentile 35-50*	0.190	-0.065	-0.154	0.040
	(0.171)	(0.132)	(0.157)	(0.126)
× *Percentile 50-65*	0.074	-0.130	-0.131	-0.052
	(0.193)	(0.137)	(0.172)	(0.125)
× *Percentile 65-80*	0.006	-0.156	-0.218	-0.206
	(0.200)	(0.126)	(0.171)	(0.116)
× *Percentile 80-95*	0.005	-0.269	0.153	-0.339
	(0.140)	(0.094)**	(0.119)	(0.082)**
Controlling for old peers quality (and interactions)	Yes	Yes	Yes	Yes

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the effects of the percentage of top 5% pupils and percentage of bottom 5% pupils obtained from one single regression including both treatments. The table displays the coefficient on treatment based on new peers. Two separate regressions run for boys and girls. Number of observations for boys: 114,927, corresponding to 38,309 pupils in 2096 schools. Number of observations for girls: 117,039, corresponding to 39,013 pupils in 2133 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 10% significant. Interaction terms obtained by interacting the percentage of top 5% pupils and percentage of bottom 5% pupils (separately for old and new peers) with a dummy indicating where the pupil ranks in terms of his/her KS2 percentiles *on average across subjects*. The effect of KS2 achievement is controlled for semi-parametrically by interacting pupil KS2 percentiles with the dummies indicating his/her rank in the ability distribution. Main effects included. In Panel B ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution.

Table 8 – Impact of peer quality on KS3 attainments (percentiles), by ability and free school meal eligibility, selected sample (std.dev.≤3)

	Pupil is eligible for free school meals		Pupil is not eligible for free school meals	
	Percentage of bottom 5% pupils, new peers	*Percentage of top 5% pupils, new peers*	*Percentage of bottom 5% pupils new peers*	*Percentage of top 5% pupils, new peers*
Dependent variable is:	(1)	(2)	(3)	(4)
Panel A: Pupils of ability pooled (overall effect				
KS3 percentiles, controlling for KS2	-0.203	0.113	-0.121	0.065
	(0.063)**	(0.061)*	(0.044)**	(0.034)*
Panel B: Ability blocks defined on original KS2 percentiles				
KS3 percentiles, controlling for KS2 (baseline)	-0.165	0.123	-0.118	0.127
	(0.059)**	(0.065)*	(0.047)**	(0.042)**
× *Percentile 20-35*	-0.045	0.105	0.006	0.100
	(0.162)	(0.161)	(0.099)	(0.085)
× *Percentile 35-50*	-0.052	-0.141	-0.007	0.050
	(0.277)	(0.258)	(0.133)	(0.099)
× *Percentile 50-65*	0.209	-0.004	-0.088	-0.060
	(0.351)	(0.295)	(0.142)	(0.099)
× *Percentile 65-80*	-0.568	-0.345	-0.067	-0.101
	(0.404)	(0.282)	(0.148)	(0.094)
× *Percentile 80-95*	-0.251	0.114	0.082	-0.258
	(0.360)	(0.266)	(0.103)	(0.070)**
Controlling for old peers quality (and interactions)	Yes	Yes	Yes	Yes

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the effects of the percentage of top 5% pupils and percentage of bottom 5% pupils are obtained from one single regression including both treatments. The table displays the coefficient on treatments based on new peers. All regressions control for the quality of old peers. Two separate regressions were run for pupils eligible and non-eligible for free school meals. Number of observations for regressions including pupils eligible for free school meals only: 36,576, corresponding to 12,192 pupils; number of schools: 2020. Number of observations regressions including pupils non-eligible for free school meals only: 195,390, corresponding to 65,130 pupils; number of schools: 2194. Standard error clustered at the school level as identified in Year 7. **: at least 1% significant; *: at least 10% significant. Interaction terms obtained by interacting the percentage of top 5% pupils and percentage of bottom 5% pupils (separately for old and new peers) with a dummy indicating where the pupil ranks in terms of his/her KS2 percentiles *on average across subjects*. The effect of KS2 achievement is controlled for semi-parametrically by interacting pupil KS2 percentiles with the dummies indicating his/her rank in the ability distribution. Main effects included. Ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution.

Table 9 – Impact of peer quality on KS3 attainments (percentiles), treatments separately defined by free school meal eligibility, selected sample (std.dev.≤11.5),

	Percentage of bottom 5% pupils, new peers		*Percentage of top 5% pupils, new peers*	
	Counting pupils eligible for free school meals only	Counting pupils non-eligible for free school meals only	Counting pupils eligible for free school meals only	Counting pupils non-eligible for free school meals only
Dependent variable is:	(1)	(2)	(3)	(4)
Panel A: All pupils				
KS3 percentiles, controlling for KS2	-0.225	-0.072	0.228	0.056
	(0.063)**	(0.041)*	(0.143)	(0.028)*
Panel B: Pupils eligible for free school meals				
KS3 percentiles, controlling for KS2	-0.241	-0.147	0.153	0.089
	(0.084)**	(0.054)**	(0.229)	(0.042)*
Pupils not eligible for free school meals				
KS3 percentiles, controlling for KS2	-0.204	-0.056	0.254	0.054
	(0.066)**	(0.043)	(0.142)*	(0.028)*
Controlling for old peers quality	Yes	Yes	Yes	Yes

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the effects of the percentage of top 5% pupils and percentage of bottom 5% pupils are obtained from one single regression including both treatments. The table displays the coefficient on treatments based on new peers and computed separately for pupils eligible for free school meals and pupil non-eligible for free school meals. All regressions control for the quality of old peers. Number of observations in Panel A: 1,866,516, corresponding to 622,172 pupils; number of schools: 2,194. Number of observations in Panel B: 256,467, corresponding to 85,489 pupils; number of schools: 2191. Number of observations in Panel C: 1,610,049, corresponding to 536,683 pupils; number of schools: 2194. Standard error clustered at the school level as identified in Year 7. **: at least 1% significant; *: at least 10% significant. Controls for old peers quality also computed separately for pupils eligible for free school meals or not.

Appendix Tables

Appendix Table 1 – Transition matrix: top and bottom 5% pupils at KS2 and their percentile scores at KS3

	Bottom 25% of KS3 percentile distribution					*Top 25% of KS3 percentile distribution*					*Rest of the distribution*	*Not entered for exam*
Variable:	≤5	5-10	10-15	15-20	20-25	75-80	80-85	85-90	90-95	95+		
Pupil in top 5% in English at KS2	1.01	0.00	0.01	0.03	0.06	6.94	9.07	13.24	17.66	37.59	13.50	0.89
Pupil in bottom 5% in English at KS2	44.37	18.59	10.15	5.49	3.52	0.53	0.45	0.41	0.26	0.31	11.68	4.24
Pupil in top 5% in Maths at KS2	0.71	0.00	0.01	0.01	0.02	4.97	7.82	12.19	16.48	51.70	5.03	1.06
Pupil in bottom 5% in Maths at KS2	41.09	28.78	8.97	4.13	2.39	0.46	0.46	0.33	0.28	0.21	9.51	3.39
Pupil in top 5% in Science at KS2	0.81	0.07	0.07	0.08	0.10	5.57	9.48	12.67	16.60	40.95	12.60	1.00
Pupil in bottom 5% in Science at KS2	32.30	27.91	12.23	6.56	3.95	0.57	0.59	0.34	0.30	0.25	11.88	3.12

Note: Cells present percentages of pupils in a given percentile score range at KS3. Percentiles are computed in the national distribution separately for every cohort. 'Not entered for the exam' includes pupils not admitted to sit for the KS3 exams because deemed below the appropriate level by their teachers; students absent on the day of the exam; and students with missing information for the KS3 test scores.

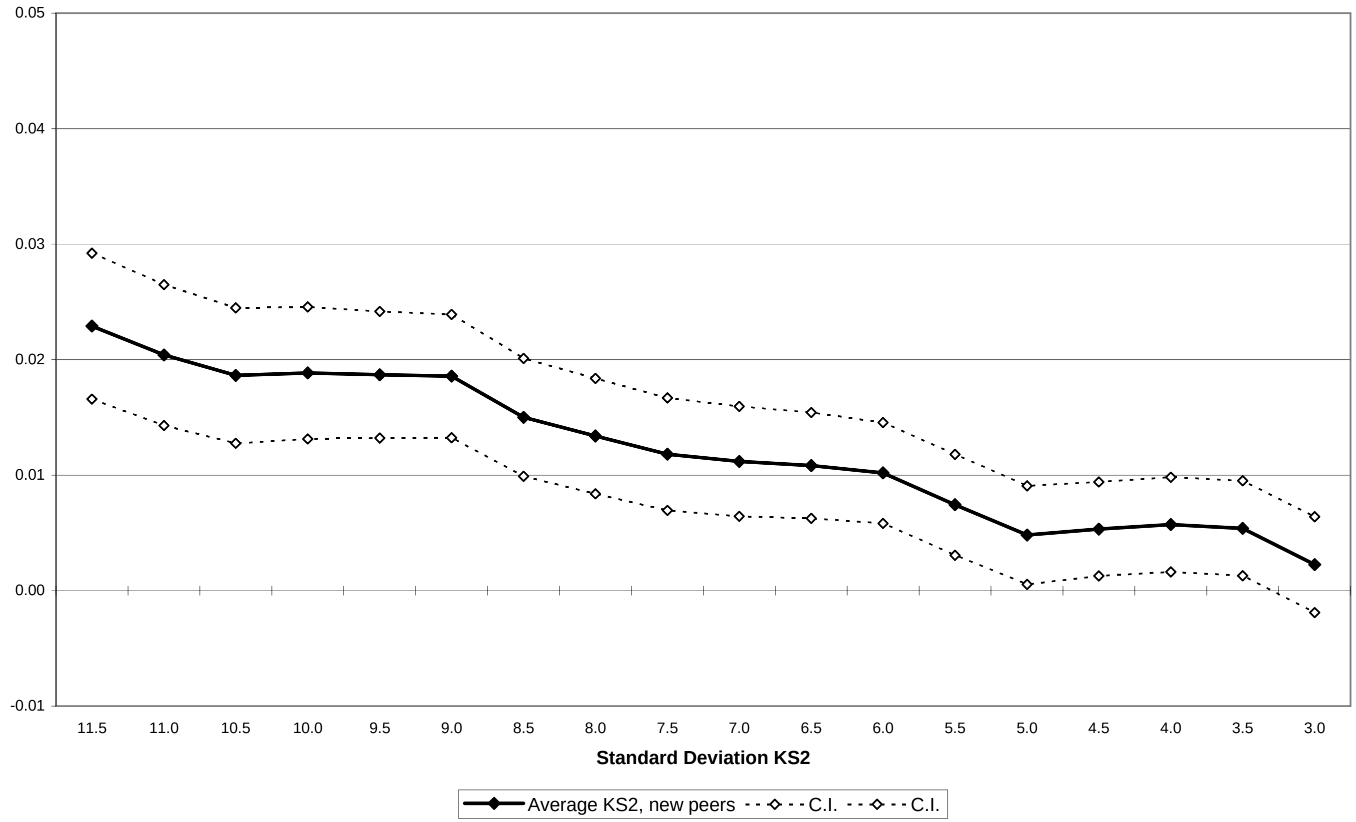
Appendix Table 2 – Additional descriptive statistics: full sample and sample with standard deviation of KS achievements (percentiles) .≤ 11.5

Variable	Full Sample	Selected Sample (std.dev.≤11.5)
Panel A: Pupils' Outcomes		
KS2 percentile, English	49.3 (24.3)	47.3 (25.6)
KS2 percentile, Mathematics	49.4 (24.3)	47.2 (25.8)
KS2 percentile, Science	48.9 (24.3)	46.8 (26.0)
KS3 percentile, English	48.9 (26.0)	47.6 (27.2)
KS3 percentile, Mathematics	49.2 (25.3)	47.4 (26.9)
KS3 percentile, Science	49.2 (25.5)	47.7 (27.3)
Panel B: Pupils' Characteristics		
First language is English	0.93 (0.25)	0.93 (0.25)
Eligible for free school meals	0.13 (0.34)	0.14 (0.34)
Male	0.50 (0.50)	0.49 (0.50)
Changed school between Year 7 and KS3	0.11 (0.31)	0.11 (0.31)
Ethnicity: White British	0.85 (0.35)	0.85 (0.35)
Ethnicity: White other	0.02 (0.12)	0.02 (0.12)
Ethnicity: Asian	0.05 (0.22)	0.05 (0.22)
Ethnicity: Black	0.03 (0.16)	0.03 (0.16)
Ethnicity: Chinese	0.00 (0.05)	0.00 (0.04)
Ethnicity: Other	0.05 (0.22)	0.05 (0.22)
Panel C: School characteristics (Year 7)		
Cohort size	201.7 (57.2)	201.9 (57.2)
Community school	0.67 (0.47)	0.67 (0.47)
Voluntary aided school	0.14 (0.352)	0.14 (0.35)
Voluntary controlled school	0.03 (0.18)	0.03 (0.18)
Foundation school	0.15 (0.36)	0.15 (0.36)
City Technology college school	0.00 (0.05)	0.00 (0.05)
Religiously affiliated school	0.16 (0.37)	0.16 (0.36)

Note: Table report means of the listed variables and standard deviation in parenthesis. Number of pupils in full sample: 1,279,514. Number of pupils in selected sample (std.dev.≤11.5): 622,172. Full sample and selected sample only include pupils with KS2 achievement in each subject above the 5^{th} percentile and below 95^{th} percentile of KS2 cohort-specific national distribution. Year 7 refers to the first year in secondary school after transition out of primary. KS3 refers to Year 9 when pupil sit for their KS3 assessment. Fractions may not sum to 1. This is due to rounding or partially missing information.

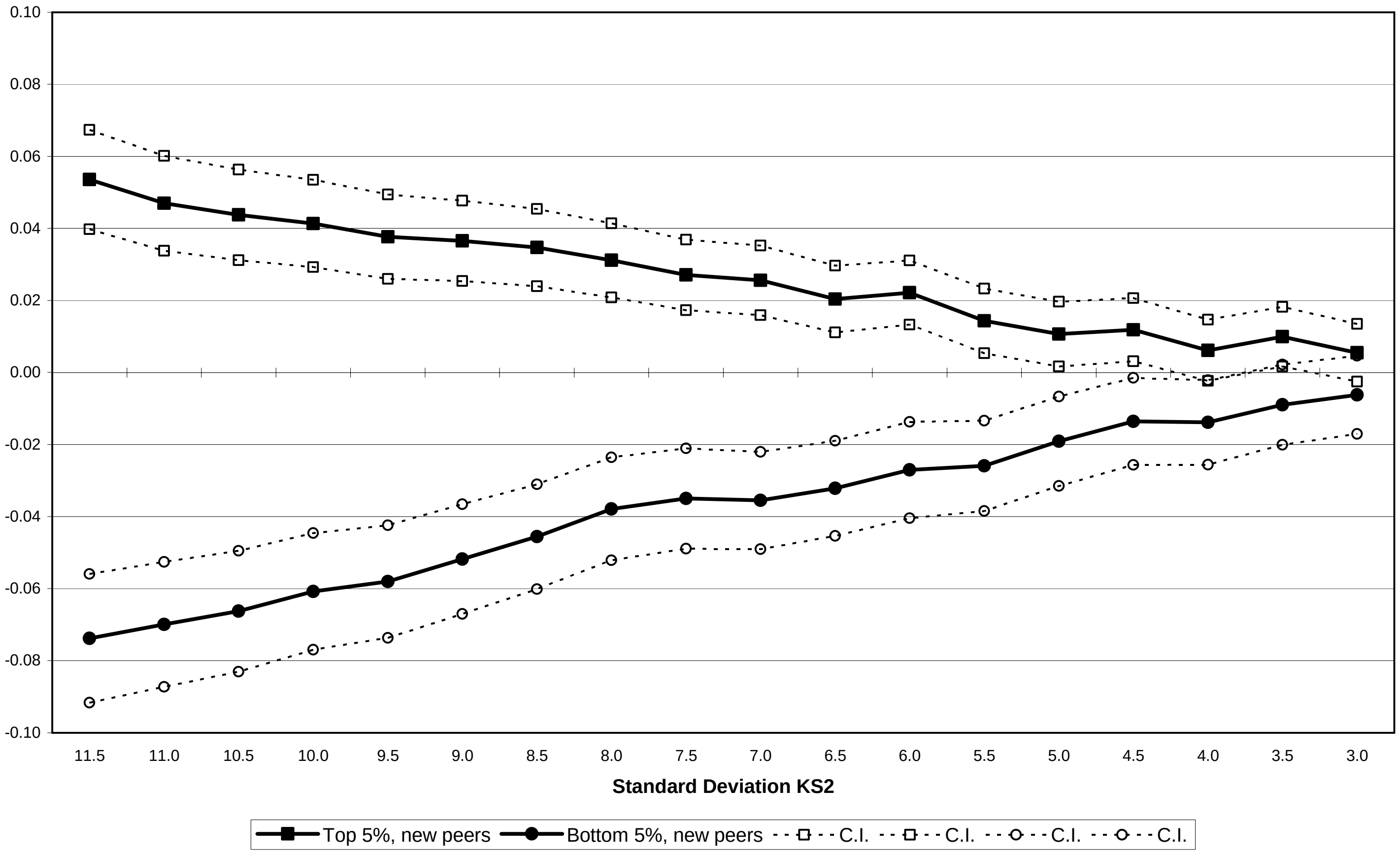
Figures

Figure 1 – Balancing of KS2 with respect to treatment, by cumulative bands of standard deviation of KS attainments (percentiles)



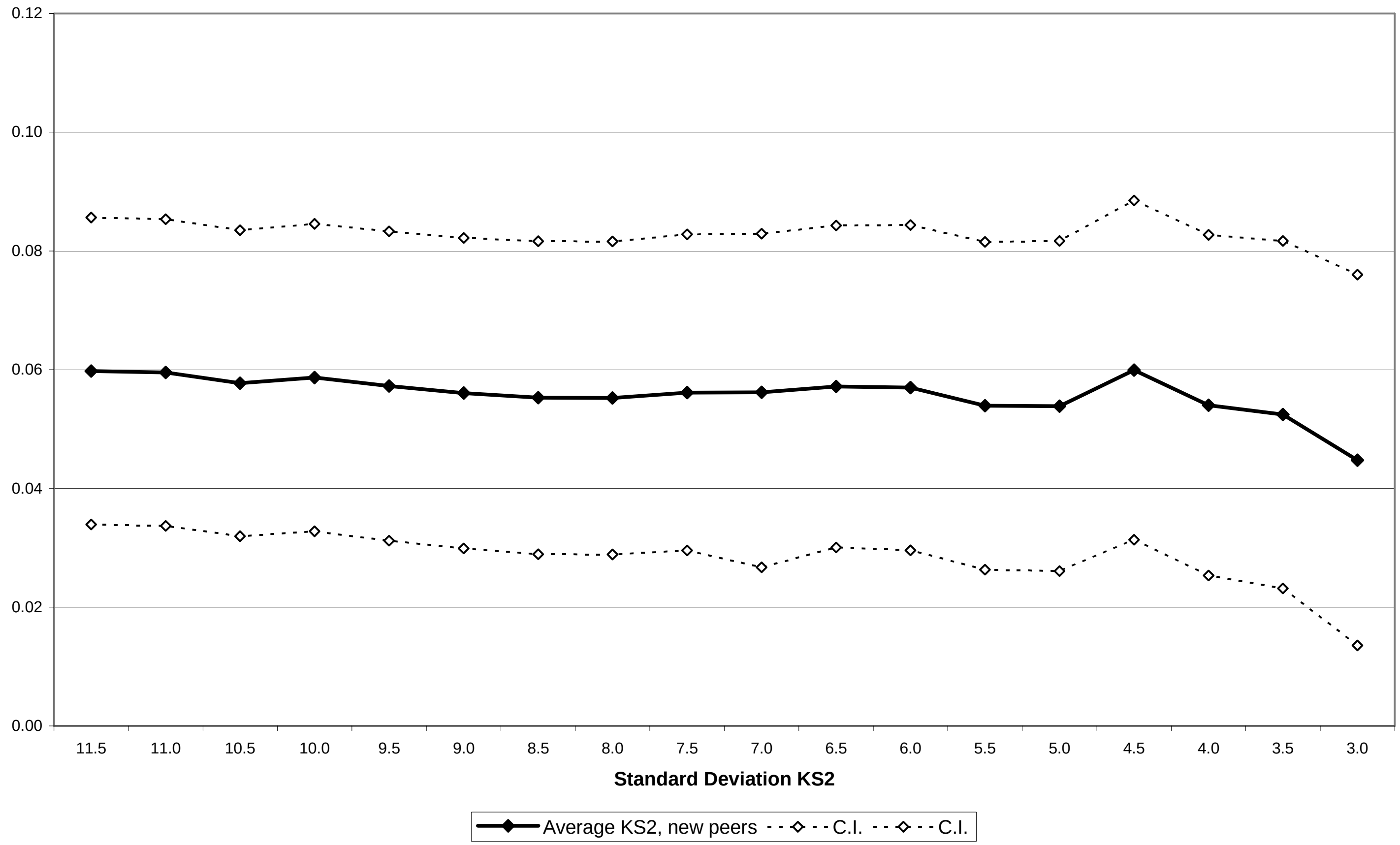
Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS2 achievements (percentiles) on the average achievement at KS2 of new peers. All regressions include pupil fixed effects and control for old peer quality. 18 different regressions were estimated over different cumulative bands of the standard deviation of KS2 attainments across subjects; these spanned the interval std.dev.≤11.5 to std.dev.≤3, in steps of 0.5. Std.dev. ≤3 includes roughly 6% of the sample; std.dev. ≤6 includes roughly 20% of the full sample; std.dev. ≤7.5 includes roughly 25% of the full sample; std.dev. ≤9 includes roughly 33% of the full sample; std.dev. ≤11.5 includes roughly 50% of the full sample.

Figure 2 – Balancing of KS2 with respect to treatments, by cumulative bands of standard deviation of KS attainments (percentiles)



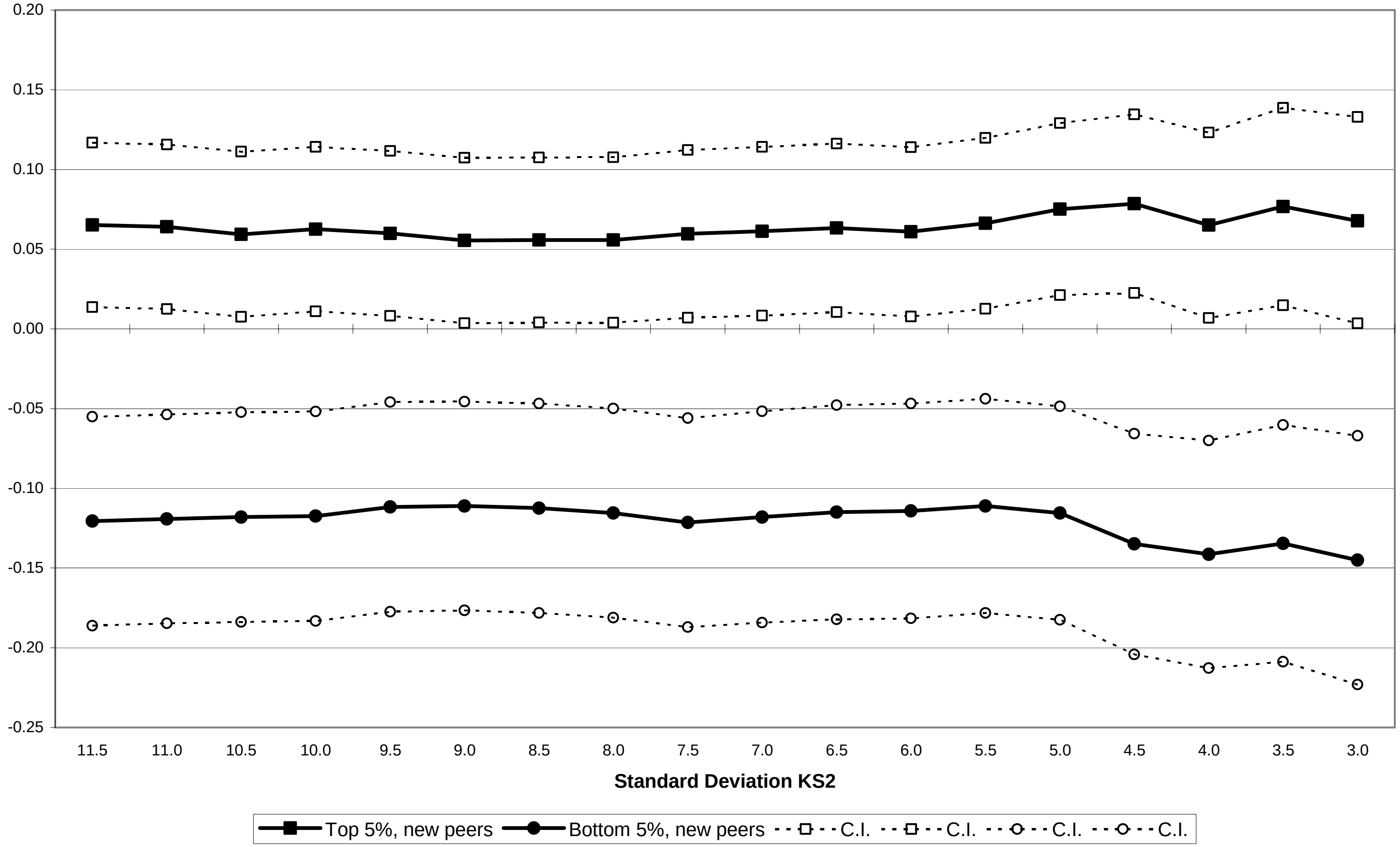
Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS2 achievements (percentiles) on the percentage of top 5% pupils, new peers, and percentage of bottom 5% pupils, new peers. All regressions include pupil fixed effects and control for old peer quality. 18 different regressions were estimated over different cumulative bands of the standard deviation of KS2 attainments across subjects; these spanned the interval std.dev.≤11.5 to std.dev.≤3, in steps of 0.5. Std.dev. ≤3 includes roughly 6% of the sample; std.dev. ≤6 includes roughly 20% of the full sample; std.dev. ≤7.5 includes roughly 25% of the full sample; std.dev. ≤9 includes roughly 33% of the full sample; std.dev. ≤11.5 includes roughly 50% of the full sample.

Figure 3 – Effect of treatments on KS3 percentiles, conditional of KS2, by cumulative bands of standard deviation of KS attainments



Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS3 achievements (percentiles) on the average achievement at KS2 of new peers. All regressions include pupil fixed effects, control for old peer quality and pupil KS2 achievement (percentiles). 18 different regressions were estimated over different cumulative bands of the standard deviation of KS2 attainments across subjects; these spanned the interval std.dev.≤11.5 to std.dev.≤3, in steps of 0.5. Std.dev. ≤3 includes roughly 6% of the sample; std.dev. ≤6 includes roughly 20% of the full sample; std.dev. ≤7.5 includes roughly 25% of the full sample; std.dev. ≤9 includes roughly 33% of the full sample; std.dev. ≤11.5 includes roughly 50% of the full sample.

Figure 4 – Effect of treatments on KS3 percentiles, conditional of KS2, by cumulative bands of standard deviation of KS attainments



Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS3 achievements (percentiles) on percentage of top 5% pupils, new peers, and the percentage of bottom 5% pupils, new peers. All regressions include pupil fixed effects, control for old peer quality and pupil KS2 achievement (percentiles). 18 different regressions were estimated over different cumulative bands of the standard deviation of KS2 attainments across subjects; these spanned the interval std.dev.≤11.5 to std.dev.≤3, in steps of 0.5. Std.dev. ≤3 includes roughly 6% of the sample; std.dev. ≤6 includes roughly 20% of the full sample; std.dev. ≤7.5 includes roughly 25% of the full sample; std.dev. ≤9 includes roughly 33% of the full sample; std.dev. ≤11.5 includes roughly 50% of the full sample.