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+

August 19, 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 these effects, we investigate which segments of the peer ability distribution drive the impact of peer quality on students' achievements. Additionally, we study which quantiles of the pupil ability distribution are affected by different measures of peer quality. To do so, we use data for all secondary schools in England for four cohorts of pupils sitting for their age-14 national tests in 2003/2004-2006/2007, and measure students' ability by their prior achievements at age-11. We base our identification strategy on within-pupil regressions that exploit variation in achievements across the three compulsory subjects (English, Mathematics and Science) tested both at age-14 and age-11. 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, and not the effect of average peer quality. Our evidence further suggests that it is the very top 5% and very bottom 5% students that matter, and not peers in other parts of the ability distribution. We also show that our results are driven by peers' academic talent, and not related to their family background. Finally, we find some interesting heterogeneity along the dimensions of pupils' ability and gender.

---

* We thank the Center for Economic Performance (CEP) at the London School of Economics (LSE) for seed money for this project. We would also like to thank Josh Angrist, Kenneth Chay, Steve Gibbons, Francis Kramarz, Steve Machin, Steve Pischke, Steve Rivkin, Yona Rubinstein, Hongliang (Henry) Zhang, and seminar participants at Brown University, CEP-LSE, CREST in Paris, EIEF in Rome, Hebrew University of Jerusalem, IFS in London and Tel Aviv University for helpful comments and discussions. All remaining errors remain our own.

‡ Hebrew University of Jerusalem, Royal Holloway University of London, 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 schools remain 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 taking 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 took their end-of-primary education national tests, which we exploit to obtain pre-determined proxy 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. The way in which we measure peer ability is a major improvement over previous studies. 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, 2009 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 to refection problems (Manski, 1993) for two reasons. 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. Secondly and crucially, we are able to track pupils during this 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 (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

[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.

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

ends of the ability distribution, namely the very 'good' and very 'bad' peers. The effect of low ability peers is significantly negative, and the effect of high ability peers significantly positive, although the former is larger and more precisely estimated. Moreover, it is mainly the very top 5% and very bottom 5% students that matter, and not peers in other parts of the ability distribution. We also document that the positive effect of 'good' peers and the negative effect of 'bad' ones are not significantly affected if we breakdown our measures of peer quality according to their family income status (proxied by free school meals eligibility). This suggests that our main findings are driven by peers' academic ability, and not related to their family background. 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 larger for pupils below the median of the ability distribution. We further explore the heterogeneity of our findings and document some interesting variation of peer effects by gender, especially regarding the impact of the very bright pupils at school.

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[4] 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 pupil's own ability in secondary school are highly correlated. This is so 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. More surprisingly, these correlations survive even when we look at the within-secondary-school variation over time of pupils' and their peers' ability (i.e. conditional on secondary school fixed-effects)[5]. This suggests that some 'sophisticated' form of sorting/selection might be taking place, with parents and schools responding to year-specific unobserved shocks to pupil and school quality. Identification strategies that rely on the randomness of peers' quality variation within-schools over time find little justification against this background.

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 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. This is perhaps unsurprising, as it simply

---

[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).

[5] A similar result is documented by Gibbons and Telhaj (2008).

suggests that pupils who excel, say, in English (relative to Mathematics and Science), will go to schools with peers who have, on average, an educational advantage in the same subject.

To overcome this form of subject-specific sorting, we follow three complementary routes. First, we add to our within-pupil specifications students' lagged test scores in a highly flexible way, such that we can control for the effect of one pupil's own ability in a specific subject, as well as his/her 'spread' of abilities across the three core-subjects. Second, we present a set of robustness checks and alternative specifications that support the causal interpretation of our findings. In particular, we include in some of our specifications school × subject fixed-effects to control for the sorting of pupils and their peers into schools based on subject-specific school unobservable attributes. Finally, as an alternative to our lagged test-scores 'control approach', we take advantage of a sub-samples of pupils that face large deviations in peers' ability across subjects, but small or no correlation between this variation and the within-pupil own variation in ability across subjects. These features are achieved by focusing on students who have small or no differences in their primary school (age-11) test scores across the three subjects. By definition, in this 'limit sample', there can be no correlation between the within-pupil across-subject variation in ability and the variation in peers' ability. Following Altonji et al. (2005), we argue that in our empirical set-up the amount of selection on *unobservable* subject-specific attributes tracks the amount of selection on *observable* subject-specific characteristics, in particular lagged tests scores. Thus by concentrating on the 'limit sample' of pupils with no variation in age-11 achievements across the three core subjects, we are able to focus on a set of pupils where the extent of sorting on unobservables should be mitigated, and to 'bound' our estimates of the causal impact of peers quality.
When we follow this strategy, we still document significant effects of secondary school peers' ability on pupils' own test scores at age-14. In fact, the causal relationships that we estimate between pupil performance in a specific subject and peers' quality in that same subject in the full sample are very similar in size and precision to the estimates we obtain in our 'limit sample'. Additionally, we can go on to replicate our estimates of the effects of peer quality for a variety of sub-sample where we allow increasing amounts of variation in students' own age-11 achievements across subjects, and thus an increasing correlation between pupils' own subject-specific ability and that of their peers. Remarkably, irrespective of the extent of within-pupil own variation in ability across subjects, we find peer effects estimates that are similar to those obtained in both the 'limit sample' and the full sample. Thus, we conclude that the observed correlation between the within-pupil heterogeneity across subjects and the variation of schoolmates' quality across subjects – as well as any related sorting on subject-specific unobservables – must be too small to confound the treatment effect of peers' ability once the estimation is based on within-student variation. Moreover, the stable pattern of estimates that we document across samples with varying degree of selection on observables suggests that selection on unobservales should have a very particular and peculiar shape in order to drive our results, for example, be totally unrelated or negatively related to the degree of selection on observables.

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 describes the institutional background and our dataset. Section 5 reports our main estimates and robustness checks, while Section 6 presents some heterogeneity in our findings. Finally, Section 7 provides some concluding remarks.

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 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 had to by-pass a variety of 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 a pupil's peer group is evidently self-selected and hence the quality of peers is not exogenous to pupil's own quality and characteristics.[6] Failing to control for all observable and unobservable factors that determine individual sorting and achievements 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; De Giorgi et al., 2009), within-school random variation (Hoxby, 2000; Hanushek et al., 2003; Ammermualler and Pischke; 2009), instrumental variables (Goux and Maurin, 2007) or sub-group re-assignments (Sanbonmatsu et al., 2004).[7] Only recently, Lavy and Schlosser (2008), 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 interactions could exert their effects. 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 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. However, it has to be

---

[6] 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.

[7] 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).

noted that the primary school setting in Kenya might not be fully comparable to more developed countries. 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.

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 the best of 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 in schools.[8] Additionally, as already mentioned, by focusing on a sub-set of pupils with little variation in prior achievements measured at the end of primary school, we identify a 'limit sample' that further helps us reducing 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. General identification strategy: within-pupil regressions

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 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 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. Ammermualler and Pischke, 2009; Hoxby, 2000; Gibbons and Telhaj, 2008; Gould et al., 2004; Lavy et al., 2008; and Lavy and Schlosser, 2007). This method potentially avoids both sources of confounding factors, 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

[8] Lavy (2009) uses the same approach to investigate the effect of instructional time on pupil academic achievements. Bandiera et al. (2008) use within-student across-subjects variation to study the effect of class size at university. Finally, Bandiera et al. (forthcoming) use within-worker over-time variation to analyse social incentives in the workplace.

investigate if this is systematically associated with the subject-to-subject variation in peers' ability. The ability peer effects that we study here are therefore 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 and four cohorts of 9th graders sitting for their age-14 national tests in the academic years 2003/2004-2006/2007, we estimate the following pupil fixed-effect equation:

$$A_{iqst} = \alpha_i + \beta_q + \gamma_{st} + \delta P_{qst} + \varepsilon_{iqst} \quad (1)$$

where *i* denotes pupils, *q* denotes subjects (English, Mathematics and Science), *s* denotes schools and *t* denotes pupils' cohort. A_{iqst} is an achievement measure for student *i* in subject *q* at school *s* in cohort *t*. 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_{st} is a school × cohort effect. Further, P_{qst} captures the average ability of peers in subject *q* in secondary school *s* in cohort *t* as measured by 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_{iqst} 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.

One advantage of this specification is that pupil fixed-effects – capturing his/her unobservable average ability across subjects and unmeasured family background characteristics– and school × cohort specific fixed-effects are 'absorbed' and fully taken care of. This is important since, as highlighted in the Introduction, we find a significant correlation between pupils' characteristics and ability, and the characteristics and ability of their peers even conditional on secondary school fixed-effects. This suggests that some form of parental sorting based on school × year specific considerations might be taking place.

Before moving on, two remarks are worth being made. First, one necessary assumption for our identification strategy is that peer effects are the same for all three subjects; stated differently, we cannot interact δ with β_q in Equation (1). Although this restriction does not seem untenable, in the analysis that follows we will provide some evidence to support this conjecture. Second, our peer effects are 'net' measures of peer influences, that is net of ability spillovers across subjects (e.g. peers' ability in English might influence pupils' test scores in Mathematics). If spillovers are very strong such that subject-specific abilities do not matter, then we are bound to find zero peer effects.

As discussed above, we are also interested in finding out which segments of the peer ability distribution are driving the average ability peer effect that we will 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:

$$A_{iqst} = \alpha_i + \beta_q + \gamma_{st} + \delta_1 P_{qst} + \delta_2 P_{qst}^{h} + \delta_3 P_{qst}^{l} + \varepsilon_{iqst} \qquad (2)$$

where δ_2 measures the effect of the proportion of peers at school 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 at school who are in the bottom 5%. Lastly, the parameter δ_1 measures the effect of the average ability of all other peers.

3.2. *Dealing with subject-specific pupil sorting and school selection*

Although the strategy described so far allows us to effectively control for pupils' average ability across subjects and school-by-cohort specific unobservables, this setup does not preclude the possibility that selection and sorting of students in different schools is partly based on subject-specific abilities. In particular, as we noted from the outset, there is a significant though small degree of correlation between the within-student across-subject variation in age-11 achievements and the variation in peers' ability across subjects.

Our first approach to account for such sorting is to control for pupils' KS2 test scores in all subjects in the within-pupil estimation. The identifying assumption here is that the lagged test scores effectively capture any unobserved ability in each subject and therefore within-subject peer assignment is as good as random conditional on primary school test scores; stated differently: there is no sorting based on other unobserved factors that are not correlated with KS2 scores. To our advantage, we can control for lagged test scores in a very flexible way by including in our specification at the same time *same*-subject lagged test scores (e.g. looking at KS3 English test score for pupil *i* controlling for his/her age-11 English achievement), as well as *cross*-subject test scores (e.g. looking at pupil *i*'s age-14 English test score controlling for his/her age-11 attainments in Mathematics). This allows us to partial out the effect of one pupil's own ability in a specific subject, as well as his or her 'spread' of ability across the three core-subjects and any cross-subject effects. Additionally, we can interact lagged test scores with subject-specific dummies, so that age-11 achievements can exhibit different effects on age-14 outcomes in different subjects. Under our most flexible (and preferred) specification, we estimate the following model:

$$A_{iqst} = \alpha_i + \beta_q + \gamma_{st} + \delta_1 P_{qst} + \delta P_{qst}^{h} + \delta P_{qst}^{l} + \lambda_q a_{iqst} + \theta_q a_{iq(-1)st} + \kappa_q a_{iq(-2)st} + \varepsilon_{iqst} \qquad (3)$$

where now a_{iqst} represents *same*-subject lagged test scores, $a_{iq(-1)st}$ and $a_{iq(-2)st}$ are the two *cross*-subjects lagged test scores, and λ_q, θ_q and κ_q are subject-specific parameters that capture the effects of lagged test scores in the same- and cross-subjects. [9,10]

[9] Note that conditional on pupil fixed effects, the same-subject and two cross-subjects lagged test scores cannot be simultaneously identified. Therefore, in our within-pupil empirical specification, we only include the same-subject lagged test score and one of the two cross-subject lagged outcome.

[10] Note that we also tried specifications where we interact pupil characteristics (e.g. ethnicity, eligibility for free school meals and gender) with subject specific dummies, and found qualitatively similar results. However, we prefer the more parsimonious specification in Equation (3).

We further complement this strategy by providing a set of robustness checks and alternative specifications that allow us to gauge the importance of subject-specific school selection and pupil sorting. In particular, we include in some of our specifications school × subject fixed-effects to control for the sorting of pupils and their peers into schools based on subject-specific school unobservable attributes.

Our last approach to control more carefully for pupil sorting based on subject-specific KS2 test scores and/or their correlates (even though unobservable) is to try to identify a sub-sample of pupils for which there are 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 topics. Mechanically, in this 'limit sample', there can not be any correlation between one pupil's subject-specific observed ability and that of his/her peers'. This is because the within-pupil variation of age-11 test scores across subjects is limited to be very close to zero (in our empirical application, the within-pupil standard deviation of KS2 test scores across the three subjects is set be at most three). Following the reasoning in Altonji et al. (2005), we argue that in our empirical set-up the amount of sorting on *unobservable* subject-specific attributes should track the amount of selection on *observable* subject-specific characteristics, in particular lagged tests scores. Therefore, by focusing on the 'limit sample' of pupils with little or no variation in age-11 achievements across the three core subjects, we identify a set of students where the extent of sorting on unobservables should be strongly mitigated.[11] In fact, the particular pattern of results that we will present below suggests that the effect of unobsevables is not driving our results.

Importantly, in our analysis we will compare estimates based on this 'limit sample' to the results that we obtain from the full sample and other intermediate samples that exhibit some degree of correlation between the within-pupil and the within-peers variation in subject-specific abilities. To preview our results, one remarkable finding is that our estimates of the ability peer effects do not change as we stretch our sample to include pupils that display progressively more subject-specific observable sorting. In other words, our strategy that controls for a student fixed-effect and lagged test scores in a highly flexible way seems robust to sorting/selection and omitted variable biases even in samples that allow for high within-pupil across-subject variation in KS2 test scores.

3.3. Measuring 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 quality 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 (6th grade) to students' KS3 achievements three years later, that is 9th grade in secondary school. Additionally, by following individuals over time, we are able to point out which secondary school students come from the same primary and identify who the new peers and the old peers are. On average, about 87% of pupil *i*'s peers in secondary school did not

---

[11] Note that identifying the 'limit sample' by imposing a restriction on the variation in lagged test scores within-pupil is analogue to within-pupil non-parametric 'matching' based on the three lagged test scores observed for each student. That is we match within-pupil on a_{iqst}, $a_{iq(-1)st}$ and $a_{iq(-2)st}$ in Equation (3), and only keep pupils with a 'close-enough' match to themselves across subjects.

attend the same primary institution as student *i*, and therefore their KS2 test scores could not have been affected by this pupil. 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. These help us to control for primary-school × cohort × subject effects that might persist on age-14 test scores and that are shared by pupils coming from the same primary school and cohort. Note however our estimates are not sensitive to the inclusion of these variables.

Before moving on two final 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 based on 7^{th}-grade enrollment. This is because any later definition of 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 school authorities may have some discretion in placing students in different classes within a grade. As a result, our estimates are more properly interpreted as 'intention-to-treat' peer effects.

4. Institutions, data and descriptive statistics

4.1. Schooling in England: institutional background

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 (KS1), spanning ages 5-6 and 6-7 (these would correspond to the 1^{st} and 2^{nd} grade in other educational system, e.g. in the US). At age 7-8 pupils move to KS2, sometimes – but not usually – with a change of school. At the end of KS2, when they are 10-11 (6^{th} grade), children leave the primary phase and go on to secondary school where they progress through KS3 (7^{th} to 9^{th} grade) and KS4 (10^{th} to 12^{th} grade). Importantly, the vast majority of pupils changes schools on transition from primary to secondary education, and move on to the school of their choice.

Indeed, since the Education Reform Act of 1988, the 'choice model' of school provision has been progressively extended in the state-school system in England (Glennerster, 1991). In this setting, pupils can attend any under-subscribed school regardless of where they live and parental preference is the deciding factor. All Local Education Authorities (LEAs) and schools must organize their admissions arrangements in accordance with the current statutory Governmental Admissions Code of Practice. The guiding principle of this document is that parental choice should be the first consideration when ranking applications to schools. However, if the number of applicants exceeds the number of available places, other criteria which are not discriminatory, do not involve selection by ability and can be clearly assessed by parents, can be used to prioritize applicants. These vary in detail, but preference is usually given first to children with special educational needs, next to children with siblings in the school and to those children who live closest. For Faith schools, regular attendance at local designated churches or other expressions of religious commitment is foremost. As a result, although choice is the guiding principle that schools should use to rank pupils' applications, it has long been suspected that they have some leeway to pursue some forms of covert selection based on parental and pupil characteristics that are correlated with pupil ability (see West and Hind, 2003).

As for testing, 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 are derived. We will use these test scores to measure pupils' attainments at KS3 and identify the quality of their peers as measured by their KS2.

Finally, regarding the organization of teaching and class formation, one important aspect that characterizes English secondary schools is the practice of 'ability setting'. Under these arrangements, secondary school pupils are initially taught in mixed-ability groups for an observation and acclimatization period of up to a year, and then educated in different groups for different subjects according to their aptitude in that specific topic. Subject-specific ability is often gauged using end-of-primary education (KS2) test scores; these are only available to schools several months after they have admitted pupils. However, teachers and school staff have some discretion in determining the ability set that is most appropriate for their students in different subjects (see DfES, 2006; Kutnick et al., 2006). Note that despite some explicit support from the Government, the practice of ability setting has not been fully adopted by secondary schools in England. Kutnick et al. (2005) reports that about 80% of secondary schools have ability sets for Mathematics between 7th grade and 9th grade, but only 53% from grade 7. These figures are much lower for English and Science respectively at: 46% (7th to 9th grade) and 34% (from 7th grade); and 59% (7th to 9th grade) and 44% (from 7th grade). In conclusion, two important features emerge from this discussion. First, ability grouping in secondary schools is subject-specific, which means that students predominantly interact with different peers in different subjects depending on their relative abilities. Second, ability setting is not strictly implemented, which suggests that pupils will face a variety of class-mates with a heterogeneous range of abilities during instruction time even for the same subject.

4.2. Data construction

The UK's Department for Children, Schools and Families (DCSF) collects a variety of data on all pupils and all schools in state education[12]. This is because the pupil assessment system is used to publish school performance tables and because information on pupil numbers and pupil/school characteristics is necessary for administrative purposes – in particular to determine funding. 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

[12] 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.

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 tests. The PLASC is integrated with the pupil's assessment records 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).

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 (6^{th} grade) through to age-14 (9^{th} grade), and to join this information to the PLASC data for every year of secondary schooling (7^{th} to 9^{th} grade). The four cohorts that we use include pupils who finished primary education in the academic years 2000/2001 to 2003/2004, entered secondary school in 2001/2002 to 2004/2005, and sat for their KS3 exams in 2003/2004 to 2006/2007. We use information on these four cohorts as our core dataset because this is the only time window where we can identify the secondary school where pupils *start* their secondary education, and not only the one where they take their KS3 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 7^{th} grade), and not after two years (in 9^{th} grade). The data also allows us to gather information about the primary school where pupils took the KS2 exams, which implies that we are able to single-out secondary schoolmates that are new peers from those who instead came from the same primary school (i.e. 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 use the KS2 test scores, separately by subject and cohort, to assign each pupil to a percentile in the cohort-specific and subject-specific national distribution of KS2. We then go on to create three separate measures of peer quality. First, we compute the average attainments of peers in the *grade* at school. Next, we create two measures that are meant to capture peer effects coming from very bright and very worst students at school, namely: the fraction of peers (in the *grade* at school) below the 5^{th} percentile or above the 95^{th} percentile of the cohort-specific national distribution of KS2 test scores.

Before moving on to some descriptive statistics, it is useful 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. 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 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.[13]

Furthermore, we apply some restriction based on the fraction of bottom 5% and top 5% pupils, in order to exclude schools with particularly high or low shares of 'good' and 'bad' peers. In particular, we drop schools where the fractions of pupils below the 5th percentile or above the 95th percentile of the cohort-specific KS2 national distribution exceeds 20%, and schools that do not have any variation over the four years in these fractions. This last restriction predominantly trims schools that have no students in either the top or bottom 5% of the ability distribution in any year and 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.

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 our core sample.

4.3. Some descriptive statistics

In Table 1 we present descriptive statistics for the main variables of interest for the sample of 'regular' students, defined as pupils with age-11 test scores in the three core subjects above the 5th percentile and below the 95th percentile of KS2 test score distribution (Column 1). The regression analysis that follows predominantly consider these pupils, which we sometimes refer to as 'treated' students. In the same table, we also presents descriptive statistics for pupils in either the top 5% or bottom 5% tails of the ability distribution, that is 'good' and 'bad' peers (which we also label as 'treatments').

In the top panel of the Table we describe pupils' test scores at KS2 and KS3. Unsurprisingly, the first column shows that for regular students test score percentiles are centered just below 50, for all subjects and at both Key Stages. The correlations of pupils' KS2 test scores across subjects are 0.60 for English and Mathematics; 0.63 for English and Science; and 0.68 for Science and Mathematics. At KS3 these correlations increase to 0.64, 0.68 and 0.80, respectively. Appendix Table 1 further shows that the within-pupil variations of the KS2 and KS3 test scores across the three subjects are respectively 11.9 and 11.2. Overall, this provides evidence that test scores are not perfectly correlated across subjects for the same student, although they tend to be more closely associated in Science and Mathematics, in particular at KS3.

The remaining two columns of the table illustrate 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. We get a very similar picture if we look at pupils' KS2 test scores in one subject (e.g. English) imposing that at least one of the other two subjects (e.g. Mathematics

[13] 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).

or Science) is above the 95th percentile or below the 5th percentile of the test score distribution.[14] 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 the achievements of very good and very bad peers between age-11 and age-14. To further substantiate this point, in Appendix Table 2 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' and 'bad' peers are consistently amongst the brightest and worst performers.

The second panel of Table 1 presents more information on pupil background characteristics. The figures in the first column confirm that our sample is fully representative of the population of secondary school pupils in England. On the other hand, pupils with at least one subject in the bottom 5% are less likely to have English as their first language and to be of White British ethnic origins, and are more likely to be eligible for free school meals (a proxy for family income). 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. Peer ability measures defined in terms of pupil background would therefore 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 7th 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. 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 students 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'. 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). Unsurprisingly, this is centered around 50 for all subjects and irrespective of whether we look at new peers or all peers. Average peer quality measures also display quite a wide range of variation, although this mainly capture differences across schools. However, Appendix Table 1 shows that there is also a considerable amount of within-pupil, across subject dispersion in average peer age-11 test scores. This is the variation that our pupil fixed-effect regressions will exploit to identify the effect of average peer quality.

In Panel B and Panel C, we present descriptive statistics for our proxies for 'good' and 'bad' peers, separately for all peers and new peers. By construction, the fractions of top 5% and bottom 5% 'new peers' in

[14] For example, the KS2 percentiles in English for pupils with at least Mathematics or Science in the top 5% and bottom 5% are 83.8 and 9.8, respectively.

the incoming cohort are smaller than the respective fractions including all peers. Once more, the wide dispersion in the fraction of 'good' and 'bad' peers predominantly picks up differences across schools. Nevertheless, Appendix Table 1 shows that the same pupil will face considerably different fractions of academically bright and weak students across different subjects. Our within-pupil specifications go on to use this variation to estimate the effect of very 'good' and 'bad' peers on pupils' own test scores at age 14.

5. Results

5.1. Effects of peers' ability: main results

We begin our discussion of the results by presenting estimates of the impact of the peer quality on pupil outcomes at KS3 obtained using the full sample of pupils and controlling for potential subject-specific sorting by including lagged test scores. Results are reported Table 3. 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. The estimates presented in the five rows of the table come from a variety of specifications, which differ in the way they control for lagged test scores. In the first row, we report estimates unconditional on age-11 achievements, while the second row presents estimates where we simply include pupils' own KS2 attainment. Next, in the third row, we include pupils' own KS2 test scores in the *same*-subject and *cross*-subject (as detailed in Section 3.2) to control for pupils' own subject-specific ability, as well as his/her 'spread' of abilities across subjects and cross-subject spillovers. In fourth row, we further interact lagged test scores in the same- and cross-subject with subject specific dummies. This allows pupils' lagged outcomes to affect age-14 test scores differently in different subjects. Finally, in the fifth row of the table we present results where we further control for KS1/age-7 pupil achievements. Primary school students are only tested in English and Mathematics at KS1, so we impute the missing test score for Science as the average between English and Mathematics. Note also that age-7 test scores have only become available recently, so we loose one of our four cohorts when controlling for KS1 test scores.

Starting from the first row, the estimates in Columns (1) and (2) suggest that average peer quality has a positive effect on students' KS3 achievements, although the within-pupil estimate is one fourth of the OLS estimate, at 0.094 versus 0.366, respectively. This suggests that controlling for pupil average unobservable ability across the three subjects by including pupil fixed-effects is a significant step in the direction of identifying the causal impact of peer quality. However, another reason why the estimates of within-pupil peer effects might be smaller is because they net out overall effects that arise through cross-subject interactions. Nevertheless, the pupil fixed-effects estimates in the first row are unconditional of KS2 achievements and thus potentially contaminated by subject-specific pupil sorting. Therefore, in the second row, we go on to include lagged test scores as an attempt to control for pupil subject-specific ability and sorting. The OLS estimate of the effect of average peers' quality is now about 20% smaller than before, at 0.305. More importantly, the within-pupil estimate is about 40% smaller at 0.058.

These results highlights two important facts. First, we find that even when we control for lagged test scores in the OLS specification, we are unable to reduce our estimate of the effect of average peer quality to values close to the within-pupils estimate. This strongly speaks in favor of within-pupil regressions, which allow us to control non-parametrically for pupils' unobservable average ability across subjects, as well as school × cohort unobservable attributes that might determine pupils' school choice. Secondly, our results reveal that, even conditional on pupil fixed-effects, there is a substantial correlation between the within-pupil variation in KS2 test scores across subjects and that of his/her peers. This is reflected in the fact that our within-pupil peer effect estimate is significantly reduced when we control for lagged test scores. We will return to this point later when we discuss the robustness of our results.

In the remaining three rows of Table 3 we go on to include lagged test scores in a more flexible way. Adding same- and cross-subject KS2 outcomes, directly or in interaction with subject specific dummies, further reduces our OLS estimate to about 0.28. More noticeably, the within-pupil estimate of the effect of average peers' quality remains stable at 0.058. It is interesting to note that when we control only for cross-subject KS2 test scores, and not for the same-subject lagged scores (results not tabulated), the average peer effect estimate is almost identical to the estimates *unconditional* on KS2 outcomes: the treatment effect estimate unconditional on KS2 is 0.094, whereas the one conditional on either one of the cross-KS2 test scores only is 0.086. This suggests that even though the within-pupil correlation among KS2 test scores in the different subjects are high, conditional on pupil fixed-effects the cross-subjects lagged test scores do not add much explanatory power to the within-pupil variation in KS3, whereas the same-subject KS2 score does. This result is an indication that the KS2 scores are not just transitory or noisy measures of subject specific ability.

Finally, inclusion of KS1 results in the last row of the table reduces our OLS estimate to 0.264, but does not significantly alter the within-pupil results now reading 0.056. All in all, this suggests that OLS estimates are quite sensitive to the way in which we use lagged test score to partial out pupils' own subject-specific academic ability. On the other hand, our within-specifications are not sensitive to how exactly we control for lagged test scores. This lends some support to our claim that, conditional on pupil fixed effects *and* controls for prior attainments, we are able to partial out pupils' unobserved overall and subject-specific abilities and identify the causal effect of average peers' quality.

As for the size of the estimated peer effect, a 10 percentile increase in the average peer quality would increase pupil own KS3 test scores by 0.58 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 worst 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 3 percentiles in the national distribution. This gain amounts to 0.12 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 0.05 of a standard deviation in the within-pupil KS3 distribution and a change from the worst to the best peer environment amounts to 0.26 of the within standard deviation of the KS3 score 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. These estimates imply that the ability peer effects that we estimate here for a 10 percentile increase in average peer quality is equivalent to the effect of half an hour of weekly instruction time in school. Another possible comparison is to the effect size of peer quality estimated in Ammermueller and Pischke (2009) 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. As for England, Gibbons and Telhaj (2009) show that one standard deviation change in average peer quality is linked to a small positive, but insignificant improvement in the test score distribution of 9th graders. Finally, Bandiera et al. (2008) study class size effects at university using a within-pupil specification similar to ours. Their results show that a one standard deviation of the within-pupil class size distribution improves test scores by 0.108 of the within-pupil standard deviation of outcomes. All in all, in comparison to studies that focus on other school inputs and interventions, our estimates capture a medium-to-small sized effect.

In the next four columns of Table 3 we test whether 'good' 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 bottom 5% new peers, while the estimates in Columns (5) and (6) report the effect of the percentage of 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. As for Columns (1) and (2), the five rows of the table differ in the way we control for lagged test scores.

The results clearly show that having more students from the bottom 5% of the ability distribution 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. However, the impact of facing a high fraction of poor achievers is larger, and more precisely estimated, than the effect of being at schools with many bright students. As for Columns (1) and (2), the OLS estimates are always much larger than the within-pupil estimates, reflecting their large selection bias. For example, in the first row, the OLS estimated impact of 'good' and 'bad' peers (0.785 and –0.959, respectively) are about five times larger that the equivalent within-pupil estimates (0.147 and –0.214, respectively). We also find that adding controls for lagged test scores significantly reduces both the OLS and the within-pupil estimates of the peer effects from the top 5% and bottom 5% schoolmates. Once again, this suggests that even conditional on pupil fixed-effects, there is a significant correlation between the within-pupil variation in KS2 test scores across subjects and that of his/her peers. Again, we return to this when we discuss robustness to school selection and subject specific ability sorting.

Moving down the remaining rows of Table 3 (Columns 3 to 6), we find that our within-pupil estimates of 'good' and 'bad' peers are not very sensitive to the exact way in which we control for KS2 test scores, and stable at around +0.061 and –0.118 respectively. This is true to a much lesser extent for the OLS estimated effects. Note also that we checked how important it is to control for the same-subject KS2 score relative to controlling for the cross-subjects KS2 scores only. The unconditional effect of the bottom 5% is –0.214, whereas the estimated effect conditional on one of the cross-KS2 subjects is –0.191. The corresponding figures for the fraction of top 5% pupils are 0.147 and 0.129. Once again, it is evident that we have to control for the same-subject KS2, but that the cross-subject KS2 test scores do not affect much our treatment estimates. This suggests that KS2 test scores appropriately pick pupils' subject-specific abilities, and not just noise and random variation in test scores.

Finally, in the last row of the Table we go on to control for age-7/KS1 test scores. When we do so, we still find a significant and negative association between the fraction of bottom 5% peers at school and pupils' own age-14 test scores in our within-pupil specification. The link between the fraction of top 5% peers and students' KS3 achievement is instead attenuated (at +0.048) and not significant at conventional levels. Note, however, that when including KS1 test scores we loose one cohort and introduce some noise in our estimates since we have to impute age-7 test scores for Science (using the average between Mathematics and English). Thus we read the results including KS1 test scores positively and providing further support to the notion that, conditional on pupil fixed effects and prior achievements, we are able to partial out pupils' overall and subject-specific abilities and identify the causal effect of peers' quality.

To conclude this section, we provide an assessment of the magnitude of the effect of these two treatments. To do so, we begin by scaling it according to the minimum and maximum values of the bottom and top treatment variables observed in the data, at zero and 20% respectively (see Table 2). A pupil who moves from 0 to 20% for the bottom quality peer environment will suffer a decline of KS3 test score of about 2.4 percentiles, which amounts to 0.09 of the standard deviation of KS3 test score, or 0.21 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 1.2 percentiles in KS3 test score, implying a change of 0.05 of a standard deviation or 0.11 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 3% and a decrease of 4% of the within-pupil standard deviation in the KS3 distribution.

We would also like to assess the effect size of our estimates relative to other studies in the field. To do so, suppose that our regular students were exposed to the following two treatments simultaneously: a reduction in the percentage of top 5% and bottom 5% new peers from 20% (the maximum in our data) to zero (the minimum in our data). This change can be viewed as a move towards class homogeneity in terms of ability, that is a sort of tracking. This shift will improve regular students' KS3 achievements by about 0.04 of a standard deviation (0.09–0.05), or 0.10 of a standard deviation (0.21–0.11) if we consider the within-pupil dispersion of KS3 achievements. Interestingly, this effect is not dissimilar from the findings in 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.

Before moving on, we notice that, contrary to Lavy and Schlosser (2008) and Lavy et al. (2008), it is difficult for us to say exactly how the peer effects that we document here might emerge. Given the English schooling institutional background and the fact that ability setting by subject is *not* strictly implemented (see Section 4), we speculate that peer effects could work both *directly*, through the interaction of students of different abilities during instructional time, and *indirectly* via the teaching body, e.g. through instructors' motivation, exhaustion and attention paid to the students with specific needs (e.g. the weakest).

5.2. Robustness of main results

5.2.1 Robustness to potential subject-specific school selectivity and pupil sorting

As discussed above, although the within-pupil specifications allow us to effectively control for pupils' average ability and school × cohort unobservables, this setup does not preclude the possibility that selection and sorting of students into schools is partly based on subject-specific abilities and unobservable features of secondary schools. In this section, we present a set of robustness checks and alternative specifications that support the causal interpretation of our previous findings. The results from these exercises are presented in Table 4. Throughout the table, Columns (1) and (2) refer to average peer quality, whereas Columns (3) to (6) come from specifications that include the fraction of peers in the top 5% and bottom 5% of the ability distribution. All specifications control for same- and cross-subject KS2 test scores, interacted with subject specific dummies, as well as for old peers' quality. Further details are provided in the note to the table.

As discussed in Section 4, parental choice is the guiding principle that education authorities should adopt when ranking pupils' application to schools. However, we have emphasized that some forms of covert selection might still take place, based on pupil and family characteristics that are associated to students' academic ability, overall or in a specific subject. Such case might arise for example for pupils attending 'specialist' schools, i.e. schools with a stated specialism in a given subject. This is because specialist schools are allowed to introduce admissions priority rules for up to 10% of their intake for pupils who demonstrate a particular aptitude in the subject of their expertise. In our sample, about 8.5% of the students attend a specialist school. Some common areas of specialism include: language; mathematics and computing; science; technology; and business and enterprise. In the first row of Table 4, we present estimates of the effects of the three measures of peers' quality obtained excluding from the sample pupils in specialist schools. These within-pupil estimates are largely comparable to those discussed in Table 3 for all peer quality measures, and if anything slightly larger than before.

Next, in the second and third row, we further look into whether results are different for non-specialist schools that are above capacity (over-subscribed) or not (at capacity or under-subscribed) on average over the four years of our analysis. As highlighted in Section 4, over-subscribed schools have some discretion in prioritizing pupils for admissions. The concern is that popular schools, receiving more admissions requests than they accommodate, might covertly select students with characteristics that are particularly suited to their teaching expertise and other school infrastructure *specific* to one of the three core subjects under analysis. On the other hand, we are not concerned with potential selection based on pupil overall ability, as this is fully taken care of in the within-pupil specifications. The results in Table 4 show that the estimates of the effects of peers' quality are similar for pupils in over-subscribed and non over-subscribed schools, in particular the impact of the fraction of top 5% and bottom 5% pupils (see Columns 4 and 6).

Further results (not tabulated, but available upon requests) additionally show that our findings are very similar for non-specialist secular schools and non-specialist schools with a religious affiliation. All in all, this evidence suggest that school side selection of pupils with characteristics (partly unobservable) that are potentially correlated with ability in a given subject is not driving our main results. In the remainder of this section, we thus try to understand whether parental choice of schools with an expertise in a given subject and/or school subject-specific unobservables might confound our estimates of peer effects.

We first examine whether our findings are driven by sorting of students who choose to attend a school with peers that excel in the same subject. More precisely, we run separate regressions for students who excel in subject q (say English) and go to schools where, on average over the four years of our analysis, new peers also excel in that subject; and for students who excel in subject q (say, again, English) and go to schools where, on average over the years, new peers excel in a different subject (either Mathematics or Science). We label these two groups 'sorted' and 'mixed' pupils, respectively. This exercise should help us understand whether our results are potentially driven by sorting of pupils with similar unobservables that are conducive to excellence in subject q (e.g. English) in the same school. Note that peers' excellence in a subject is defined using new peers' average KS2 test scores, or the fraction of new peers in the top 5% of the ability distribution, depending on the 'treatments' that we focus on.

Results are reported in Rows 4 and 5 of Table 4 and reveal that for both 'sorted' and 'mixed' pupils the effect of average peer quality is around 0.05. Furthermore, Columns (4) and (6) show that the effect of being at school with a larger fraction of top 5% peers is more positive for 'mixed' pupils (at 0.071) than for 'sorted' students (at 0.058), and that the negative effect of poor learners is larger for 'mixed' pupils (at –0.140) than for 'sorted' ones (–0.102). This pattern significantly helps ruling out that pupil subject-specific sorting might be driving our results. To see why, let us consider a simplified example with only two subject, English and Mathematics, and two pupils i and j, the first excelling in English and the second excelling in Mathematics. Both students go to the same school k, where new peers excel in English. Our results suggest that pupil j's English results (his/her weak subject) will improve slightly more than pupil i's English test scores (his/her strong subject), when meeting new peers that are strong in English. On the other hand, pupil j's test scores in Mathematics (his/her strong test score), will suffer more from meeting more new peers that are *not* good in that subject, than for pupil i, who is weaker in Mathematics. These results can hardly be explained by some mechanical sorting of pupils with similar subject-specific unobservables into the same schools, which might lead them to excel in a given topic and under-perform in the others.

In the last robustness check we control for school subject-specific unobservables (fixed over time), such as teachers' expertise in a given field, directly by including in the regressions school × subject fixed-effects. To start with, in Row (6) of Table 4, we drop from our specifications pupil fixed-effects and only control for school × subject fixed-effects and school × cohort fixed-effects. This specification allows us to net out school subject-specific (fixed over time) unobservable attributes – such as specialism in a given field – and school cohort-specific (constant across subject) unobservables – such as changes in head-teacher or finances available to the school, as well as cohort-specific shocks to the quality of pupil intake and teaching staff. However, by dropping pupil fixed-effects, we rely on the inclusion of lagged test scores to partial out pupils' ability (overall and subject-specific) in a parametric way. When we follow this strategy, we find that the negative effect of a large fraction of bottom 5% peers is very close to what we previously found, at –0.114. However, the impact of average peer quality and the effect of the top 5% peers are now twice as big as what we obtained using within-pupil specifications, respectively at 0.141 and 0.121. This suggests that controlling for pupil overall ability in a non-parametric way by including pupil fixed-effects brings us closer to an estimate of the 'true' causal impact of peer quality than controlling for school × subject and school × cohort fixed-effects at the expenses of relying on more parametric methods to partial out students' overall ability.[15]

Finally, in the last row of Table 4, we go on to include at the same time pupil fixed-effects and school × subject fixed-effects (school × cohort fixed-effects are 'absorbed' in the within-pupil specification). We estimate this specification using only the first and last cohort in our data in order to maximize the variation that we can exploit to estimate peer quality effects. Indeed, this approach is very demanding because, conditional on fixed effects, our data shows very little within-school-subject variation over time, in particular in terms of students' age-14 outcomes. In fact, the within-pupil variation of KS3 test scores is 11.29, and only shrinks to 10.22 if we further absorb school × subject fixed-effects. This suggests that the 'spread' of pupils' KS3 test scores around their average is not significantly widening or vanishing over time within schools, which is perhaps not surprising given that we are considering standardized test scores and schools' composition does not dramatically changes over four years. Even then, our results clearly point in the right direction. The effect of the average peer quality is estimated to be 0.014, that is about one fourth of what previously found. The results for bottom 5% and top 5% peers are more robust, respectively at –0.072 and 0.033, or about 60% and 54% of the estimates discussed above.

5.2.2. Estimates based on the 'limit sample'

In this section we explore the idea that by identifying a sub-set of pupils where sorting on observables is fully eliminated, we are likely to substantially reduce the amount of selection on unobservables and thus bound the 'true' causal effect of peer quality. This sort of arguments appears often in quasi-experimental studies, such as those based on regression discontinuity designs (e.g. Imbens and Lemieux, 2008), and has been articulated in Altonji et al. (2005). We pursue this reasoning by presenting estimates based on a 'limit sample' of pupils that face substantial variation of peers' ability across subjects, but have small or no differences in their own primary school test scores across English, Mathematics and Science. Stated differently, we sample students that might exhibit different levels of overall ability (which will be 'absorbed' by pupil fixed-effects), but whose observable subject-specific abilities are 'balanced'. The underlying assumption for this strategy to help

[15] Note that our new peers' quality measures actually vary at the primary × secondary school level, and by subject. This means that we can control for primary × secondary × subject fixed effect and partial out any form of primary-secondary sorting based on subject specific considerations. If we do so, we find results that are similar to those presented in Table 4. Results are available from the authors.

reducing the extent of sorting on unobservables and purge remaining biases from the peer quality estimates is that the amount of selection on *unobservable* subject-specific attributes tracks the amount of selection on *observable* subject-specific characteristics, in particular lagged tests scores. Put simply, the amount of unobservable and observable subject-specific sorting should be positively correlated.

We being this investigation by discussing Figures 1 and 2. The 'limit sample' includes pupils with a standard deviation of KS2 achievements across subjects smaller than or equal to three. Figure 1 plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained from regressions of pupil KS2 achievements on the average KS2 achievement of new peers. Figure 2 shows regression coefficients and confidence intervals obtained from regressions of pupils' own KS2 achievements on the percentage of top 5% new peers and percentage of the bottom 5% new peers. All regressions include pupil fixed-effects and control for old peers' quality. To obtain the figures, 23 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, and then std.dev.≤15; std.dev.≤17.5; std.dev.≤23; std.dev.≤26; and full sample. More details are provided in the notes to the figures.

Figure 1 shows that in all groups, but the 'limit sample' with s.d.≤3 there is a significant positive correlation between one pupil's own KS2 achievements and the average achievement of his/her new peers in secondary school. This is clear evidence of a 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 the sample with s.d.≤3 this correlation is no longer significantly different from zero and very small. In Table 5, Column (1), we provide more evidence on the degree of balancing that is achieved in the 'limit sample'. Starting from Panel A, OLS estimates (i.e., without pupil fixed-effects) show a strong positive correlation between student characteristics and the KS2 average 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, adding the pupil fixed effect eliminates the strong and significant positive correlation of pupils' and their peers' KS2 achievements. Indeed, the OLS estimate of this link is 0.292, whereas the within student estimate is 0.002.

Next, Figure 2 shows a remarkably symmetric convergence from high to zero correlation between the fraction of 'good' and 'bad' peers at school, and one pupil's own KS2 achievements 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 in the full sample, but it converges to almost zero and becomes insignificant at s.d. ≤3. On the other hand, 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.≤3. To further illustrate the validity of our strategy, Columns (2) and (3) of Table 5 present OLS and within-pupil evidence on the balancing of individual characteristics with respect to the top and bottom peer quality treatments for the 'limit sample'. Once again, 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.

Table 6 present estimates of peer effects obtained using the 'limit sample' of pupils with s.d.≤3. The various rows of the table differ in the way we control for pupils' lagged test scores. Unsurprisingly, since students in this sub-set have balanced KS2 outcomes across subjects, it does not make much difference whether we control for lagged test scores in the within-pupil specifications or not. On the other hand, inclusion of age-11 test scores still significantly reduces OLS estimates, highlighting the importance of controlling for pupil overall ability using fixed-effects. Column (2) shows that our estimate of the impact of average peer quality are not too dissimilar from what we found above. In the specification where we control for same- and cross-subject KS2 test scores in interaction with subject specific dummies, we find a peer effect of 0.055; this compares with an effect of 0.058 in the full sample using the same specification (fourth row of Table 3, Column 2). On the other hand, we now find slightly larger estimates for the effect of the top 5% and bottom 5% new peers, at 0.097 and –0.149 respectively in our preferred specification (Row 5). This might be partly due to the fact that the 'limit sample' that yields perfect balancing is not representative of the full sample of secondary pupils in England; descriptive statistics are provided in Appendix Table 3. As it turns out, the sample of pupils with s.d.≤3 includes a relatively large proportion of pupils with low and high KS2 achievements, although the former is bigger than the latter. As we will document later in the paper, pupils in different part of the ability distribution are affected differently by the very bright and very poor achievers, which helps explaining the discrepancy between results in the full sample and in the 'limit sample'.

However, the most remarkable finding from this exercise is that, even as we shrink the within-pupil standard deviation of KS2 test scores to be close to zero, we still find significant estimates of the effect of peer quality on pupils' own age-14 test scores. In the 'limit sample', there is mechanically no room for sorting based on pupils' and their peers' subject-specific observable abilities. If the amount of selection on *unobservables* tracks the amount of selection on *observables*, estimates based on the 'limit sample' should be close to the true causal effect of peer quality, as biases due to subject-specific unobservables should be mitigated. Even more remarkably, we find that the peer effect estimates are roughly constant for other sub-samples of pupils where we allow an increasing amount of within-pupil standard deviation of KS2 test scores and thus potentially more significant amount of subject-specific sorting. We present this evidence in the next section.

5.2.3. Effects of peers' ability in extended samples

We noted above that the 'limit sample' is not fully representative of the whole population of students in England and that this might explain the small discrepancy between our findings for this sub-set of pupils and in the full sample. To shed light on this issue, in Figures 3 and 4 we present estimates of the treatment effects alongside with confidence intervals obtained from 23 separate regressions that progressively use sub-samples of students with larger standard deviations of KS2 attainments across subjects. Note that these sub-groups are identical to those used to check balancing properties of the peer quality treatments (Figures 1 and 2). All regressions include pupil fixed-effects and control for old peer quality, as well as for pupils' KS2 own achievements.

Starting from Figure 3, one remarkable result is that the estimates obtained from the various sub-samples are not remarkably apart from one another. In particular, consider the sample of pupils with s.d.≤11.5: as shown in Appendix Table 3 pupils in this sub-set are very close to being fully representative of the population of students in English secondary schools. Figure 3 and Appendix Tables 4 show that estimates of the effect of average peer quality for this sample, at 0.066 in our preferred specification, are close to those obtained in the full sample (0.058) and in the 'limit sample' (0.055). Moreover, the confidence intervals throughout the Figure 3 are largely overlapping, allowing to reject the hypothesis that the estimates are different. This evidence clearly suggests that the imbalance in KS2 that emerges as we move to larger, less selected and more representative sub-sets of pupils (see Figure 1), is too small to confound our estimates of the effect of average peer quality on students' own KS3 test scores in a within-pupil regression.

Figure 4 and Columns (3) to (6) of Appendix Table 4 present similar evidence for the effects of 'good' and 'bad' peers. Once again, we find that the peer effects estimated from the sample with s.d.≤3 are very close to those obtained for a variety of sub-samples and for the full sample. For example, the top 5% peer effects estimated in the full sample, in the sample with s.d.≤11.5 and in the 'limit sample' are 0.064, 0.082, and 0.097, respectively (using the most flexible specification). The corresponding figures for the fraction of bottom 5% peers are –0.124, –0.126 and –0.149. Additionally, the lines connecting the point estimates of the effects of 'good' and 'bad' peers for all the other sub-samples are almost horizontal throughout Figure 4, with largely overlapping confidence intervals. All in all, this evidence reinforces our intuition that the imbalance in KS2 – and any related subject-specific unobservable – that we obtain as we move to less selected samples is sufficiently small not to confound our estimates of the effect of peer's quality conditional on pupil fixed-effects and lagged test scores controls.[16] This evidence also suggests that any bias due to confounding subject-specific unobservables must have a very special pattern so as to lead to the same or larger point estimates of peer effects in different samples with a shrinking degree of sorting on observables. In particular, selection on unobservables must be totally uncorrelated or negatively related to selection on observables – most prominently lagged test scores – in order to explain these results. This is highly implausible since we have shown that KS2 test scores are quite reliable proxies of pupils' subject-specific abilities, and since it is very likely that pupils with similar subject-specific abilities or preferences will sort in the same schools.

5.3. The good, the bad and the average: who has an effect on regular students?

To disentangle which segments of the distribution of peer ability drive the impact of the average peer quality on pupil KS3 test scores, we next 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 7 reports regression coefficients and standard errors from a variety of specifications. All regressions include pupil fixed-effects. 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, 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

[16] Note that we have carried out additional exercises to check the size and significance of our estimates based on the following mutually exclusive bands of the standard deviation of KS2: 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 found that the estimated effects of 'good', 'bad' and the average peers are virtually identical across the sub-samples. Results are not displayed, but are available from the authors.

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 of the table 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 that it is mostly the top and bottom peers that have an impact on regular students in schools, and not peers of average quality. Indeed, the estimates of the effect of the top and bottom peers are very similar to those obtained when average peer quality was not included in the specification (compare with Table 4). On the other hand, the effect of average peer quality is substantially reduced at 0.037 in our preferred specification, that is about 36% smaller than when it was included as the only proxy of peers' ability (see Table 4).

In Figure 5, we go one step further and provide evidence that not only it is the top and the bottom that mainly matters, but also that it is the very top 5% and very bottom 5% new peers that are more strongly associated with pupils' own age 14 test scores, and not the quality of peers in other parts of the ability distribution. The figure presents treatment effect estimates and associated confidence intervals for different measures of the bottom and top new peers. For the bottom treatment, we define the following five groups: bottom 5%; 5 to 10%; 10 to 15%; 15 to 20% and 20 to 25%. For the top group, we define the following five peer measures: top 5%; 90 to 95%; 85 t0 90%; 80 to 85% and 75 to 75%. The specification further include pupil fixed-effects and controls for old peer quality, average peer quality and students' own KS2 test scores; the sample of 'treated' pupils includes students in the range from 25th to the 75th percentiles of KS2 test scores.

Figure 5 reveals a marked, though asymmetric, U-shape pattern. All the five bottom peer groups have a negative effect on other pupils, but the effect is clearly significant only for the first two groups, and it is declining sharply in scale as we move away from the very bottom. On the other hand, the effect of the top peers at school is significant only when the very top 5% pupils are considered, though the estimates are positive for the other four groups as well. This suggests that our choice of top 5% and bottom 5% peers is far from arbitrary and justified by the fact that most of the positive and negative peer effects come from the outstandingly good students and the very poor learners, respectively.

5.4. Is it cognitive ability or family economic status of peers?

In this section we examine whether there is any heterogeneity in ability peer effects by peers' family economic status, as measured by pupils' eligibility for free school meals (FSM), a proxy for family income. The answer to this question helps disentangling whether our main findings are driven by the academic ability of peers or by their family background.

The results from this exercise are tabulated in Table 8 and are based on a specification where we simultaneously include four different treatments, namely: 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 (Column 3); and the fraction of top 5% peers non-eligible for FSM (Column 3). 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 proxies 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, but 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). Thus we expect the results to be 'noisy' and not very precisely estimated. Note also that we concentrate on the top 5% and bottom 5% peers, while neglecting average quality peers, since we have documented that this is where most of the empirical action lies.

The results provide clear evidence that pupils in the bottom 5% of the ability distribution have a negative and significant impact on the KS3 achievement of regular students, no matter whether these peers are eligible for FSM or not. Although, the treatment estimate of the bottom 5% peers from low income families is larger than for non-FSM peers (at –0.242 versus –0.071, respectively), these impacts need to be re-scaled to account for the higher within-pupil standard deviation of the treatment variable in Column (1). The numbers in italics at the bottom of Table 7 refer to the 'effect size' calculated as the impact of a one standard deviation of within-pupil distribution of peers as a percentage of one standard deviation of the within-pupil distribution of KS3 percentiles. These scaled effects show that the treatment effect of bottom 5% pupils from low income families is still about twice as large as the negative effect of poor achievers from better-off families, although the difference is much less then when looking at the un-scaled coefficients. One possible explanation for this finding is that poor learners in a specific subject from wealthier families might have access to private remedial tuition, thus mitigating their negative peer effect on other regular students.

On the other hand, the evidence on the positive effect of the very talented peers on other students is much more similar for the two types of treatments. Indeed, we find a positive effect of the top 5% peers irrespective of whether they are eligible for FSM or not, and the re-scaled effect of bright pupils from poor families is 0.428, which is not dissimilar from the effect of good peers from better-off families, at 0.627. Note that only the result for peers who are not eligible for FSM is significant at conventional levels. However, 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 ones are driven by their academic ability, and not predominantly related to family background.

5.5. Additional findings: peer effects estimates by subject coupled

We mentioned in Section 3 that one of the underlying assumption of the identification strategy is that peer effects are constant across different subjects. Although this assumption is difficult to test, we provide some related evidence in Appendix Table 5, where we run regressions separately for couples of subjects, that is by pooling pupils' observations for: English and Mathematics only (top row); English and Science only (second row); and Mathematics and Science only (bottom row). Results come from specifications that include the three new peers' quality measures simultaneously, and controls for old peers' quality.

The findings suggest that most of our identification comes from the comparison of English with Mathematics and English with Science. For these two couples of subjects, we find significant and similar effects of both average peer quality and the fraction of peers in the bottom 5% of the ability distribution. The positive effect of pupils in the top 5% of the KS2 distribution is only significantly estimated for the couple English-Mathematics, although even when considering English and Science we find a positive effect at 0.032. On the other hand, our results are much weakened (in particular the effect of average peer quality) when we only pair Mathematics and Science. This is perhaps unsurprising given two sets of considerations. First, as discussed in Section 4, pupils' KS3 test scores are much more correlated for Science and Mathematics (0.80), than for English and Mathematics (0.64) or English and Science (0.68). As a result, there is less within-pupil across-subject variation in age-14 test scores to precisely estimate peer quality effects. Indeed, the within-pupil variations for English-Mathematics and English-Science are 10.8 and 10.2, respectively 35% and 27.5% higher than the within-pupil variation for Mathematics-Science, at about 8.0. Second, the institutional details presented in Section 4 revealed that 'ability setting' is more common in Mathematics and Science than in English. Given the high correlation between pupil's attainments in these two subjects, it is likely that the one student will be 'set' at a similar level in these two subjects, thus facing peers of similar quality in both Science and Mathematics. Stated differently, the within-pupil variation of the peers that the student *actually* interacts with (which we do not measure as we do not have class level data) might be too small to identify a significant peer effect. All in all, however, we believe the findings presented in this section support our assumption that peer effects are similar across subjects.

6. Allowing for heterogeneous effects

6.1. Heterogeneity by students' ability

In this section, we test for the presence of heterogeneous effects along a variety of dimensions. We first examine if the very good, the very bad and the average peers differentially affect students with different academic abilities. For this purpose, we stratify the sample into five groups according to the distribution of pupils' *average* of their KS2 percentiles across subjects. The percentile-ranges that define 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 simultaneously include interaction terms of the percentages of top 5% peers, bottom 5% peers and average peer quality (separately for old and new peers) with dummies indicating to which of the six KS2 ability groups a pupil belongs to. Note that the effect of KS2 achievements in the *same*- and *cross*-subject is controlled for semi-parametrically by interacting pupils' own KS2 percentiles with the dummies indicating his/her rank in the ability distribution (main effects are included).

The findings are reported in Table 8. The estimates presented in Column 1 reveal that the quality of average peers affects regular pupils similarly across the ability distribution. For pupils in the 5^{th} to the 80^{th} percentile of the KS2 distribution this effect is estimated to be around 0.35, which is not dissimilar to what we showed in Table 7. Only for pupils in 80^{th}-95^{th} percentile interval the average peer quality effect is higher at 0.050, although an F-test clearly accepts the null that all the coefficients are the same across percentiles. Similarly, Column (2) shows no clear variation in the negative effect of the bottom 5% peers across various ability groups of regular students. The only exception is for pupils in the 80^{th} to 95^{th} percentile of the ability distribution, where the impact of a large fraction of bottom 5% peers is still estimated to be negative, but only at –0.080 and statistically insignificant. Nevertheless, an F-test on the equality of all coefficients comfortably accepts the null.

Results for the effect of peers in the top 5% of the ability distribution reveal a more interesting pattern; these are presented in Column (3). The positive effect of this treatment is seen in all groups, although for the two top ability sets of regular students this impact is much reduced. For students in the group at the 65^{th}-80^{th}

percentiles the positive effect of very bright peers is about half of what previously found, at 0.032, and not significant at conventional levels. More remarkably, the effect of the top 5% peers is virtually zero (0.009) for pupils in the 80th to 95th percentile of the KS2 distribution. Although an F-test on the equality of the estimates across ability groups still accept the null of no differences, this result is still quite remarkable. Note that we find a similar pattern when using various pupil sub-samples, such as the 'limit sample' discussed above or the sub-set of students with a within-pupil standard deviation of KS2 attainment below 11.5.

What could explain this result? One possible explanation is that this is a mechanical 'crowding-out' 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 (80th-95th percentiles) from advanced courses or activities, such as Science and Mathematics clubs or special field trips because of limited space available in such activities.[17] To clarify this, recall that there is some degree of ability setting by subject in secondary schools (see Section 4), and consider that there is usually only a limited number of places available in top-tier classes for each subject in each school (irrespective of cohort size). Under this scenario, having many good peers *in that subject*, has two 'competing' effects for regular pupils, in particular for those in the top part of the ability distribution. On the one hand, there could be a positive effect that works either *directly* through interaction of students during instructional time, or *indirectly* via the teaching body (e.g. instructors' motivation). On the other hand, a large share of outstanding peers would reduce one student's chances of getting into the top class and participating in advanced level learning, thus depressing his/her motivation and ultimately potentially harming achievement. This counter-balancing effect should be more pronounced for the next-to-the-most able students, i.e. pupils in the 80th to 95th percentile of the ability distribution.

In fact, the implication of this reasoning should vary by cohort size: in smaller schools, the positive effect of having many top 5% peers should prevail, since there is at the same time more room for interactions of pupils of different abilities and less scope for crowding-out of good students from top-tier classes. On the other hand, in larger schools the positive effects of peers must mainly work via the teaching body, as outstanding peers will be often taught in different classes and have fewer interactions with other students. However, crowding-out of next-to-the-most able students from the best classes is more likely to happen, as places in top-tier activities are more rationed given a large cohort size.

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% peers, as well as from average peer quality, are estimated to be larger and more significant in smaller schools, in particular for those in the bottom half of the cohort-size distribution. This is not surprising given that pupils' mixing and students' interactions during instructional time will be more frequent in schools with smaller cohorts.

Additionally, we find that for schools in the bottom half of the cohort-size distribution the positive impact of the top 5% peers is positive and roughly constant throughout the ability distribution of regular

[17] 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.

students. On the other hand, we find that for school in the top half of the cohort-size distribution, this effect is positive for the least able regular students, but becomes negative (insignificant) for students in the top half of the ability distribution.[18] This evidence suggests that a crowding-out explanation of our findings might bear some relevance. However, we cannot exclude other more subtle explanations often discussed in the educational and psychological literature, such as students' competitive pressure and "big-fish-small-pond" mechanisms (see Marsh, 2005). Given the information contained in our data, these alternative hypothesis must remain conjectures.

Note that another possible and rather mechanical explanation for why pupils who are good on average do not benefit from having many top 5% peers might be related to mean-reversion. In general, average test scores *do* reveal some mean reversion. For example, pupils in the 5^{th}-20^{th} percentile at KS2 experience a 4 percentile point average improvement in their average 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 of KS2 for students in the same ability group must be similar by construction. This means that all pupils within the same ability group, in particular those in the 80^{th}-95^{th} KS2 percentile, would be similarly affected by mean-reversion *irrespective* of how many good peers they interact with. Moreover, if mean reversion was to explain our findings, there is no reason to believe that 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 checked whether the pure effect of belonging to 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 no evidence that simply belonging to this ability group significantly (negatively or positively) affects KS3 outcomes, nor do we find any relation for other ability groups. 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 zero effect discussed here above. Mean reversion does not appear to be a likely explanation for our results.

6.2. Gender heterogeneity in treatment effects

The heterogeneity of peer effects by gender is also particularly interesting, especially in secondary schools where the social interactions between boys and girls intensifies. Therefore, in Table 10, we report some results based on separate samples for boys and girls.

Our results show that the effect of the bottom 5% peers is negative and significant in both gender groups, although is it smaller for boys (at –0.076) than for girls (at –0.098). On the other hand, the effect of the top 5% peers is positive, significant and sizeable at 0.091 for girls, but negative for boys at –0.052, although this estimate is only significant at the 10% level. As for average peers' quality, our estimates by gender are not very precise, although they are more positive for girls.

[18] Results are not tabulated for space reasons, but are available from the authors upon request.

To further investigate these patterns, 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 9). The results are tabulated in Panel B of Table 10. While the effect of many 'bad' peers and of the average peer quality are not remarkably heterogeneous, we note that the finding of a negative effect of top 5% peers on boys is evident for almost all ability groups, but it is larger and more precisely estimated for the most able regular students. For girls, the effect of top peers is positive for almost all ability groups, but this estimate looses significance for the most able students.

Since these results are somewhat unexpected, in particular the finding that boys are significantly negatively affected by having a high proportion of very bright students at school, we performed a series of checks to assess to robustness of these findings. Firstly, we estimated models using the 'limit sample' discussed above, where we cap the extent of sorting on observable subject-specific abilities, and found very similar patterns. 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. Finally, another possible explanation is that the negative effect is mechanical, and once more due to mean-reversion or a ceiling effect. However, this does not seem the case. Therefore, a natural conclusion is that these effects must be 'real', and the main question is whether this pattern of heterogeneity by gender can be related to other findings in the literature. Broadly speaking, the answer to this query is positive as a growing body of evidence shows that that girls are more affected than boys by education inputs and interventions.[19]

6.3. Heterogeneity in treatment effects by pupils' eligibility for free school meals

In this section, we examine heterogeneity in ability peer effects by pupils' eligibility for free school meals (FSM), a proxy for family income. Results are shown in Table 11. The structure of the table is identical to that of Table 10, 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 for all students. More interestingly, we find that pupils eligible for FSM are affected by a greater margin by both the bottom 5% peers and average peer quality, than pupils who are not eligible for FSM. Note that this is not just because the within-pupil standard deviation of the two treatments is different for the two sub-sets of pupils. The impact of the top 5% students is also slightly larger for pupils from poorer families, although this effect is not precisely estimated. This lack of

[19] 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). Some recent studies also show a consistent pattern of stronger female response to financial incentives in education, with the evidence coming from a variety of settings. Angrist and Lavy (2009), 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 effects on males (Angrist et al., 2009). 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. Finally, a number of public-sector training programs generated larger effects on women than men (Lalonde, 1995).

precision can be explained by the considerable drop in sample size when focusing on pupils who are eligible for FSM (about 13% of our regular students).

In Panel B of the table, we further break down the treatment effect estimates by pupils' own ability, and again separately according to pupils' eligibility for FSM. The results show that average peer quality and bottom 5% peers have roughly constant effects throughout the ability distribution of regular students, irrespective of one pupil's eligibility for FSM. As for the effect of 'good' peers, we find positive estimates throughout the ability distribution for pupil from better-off families. On the other hand, the effect becomes negative at the top of the ability distribution for students who are eligible for FSM, although these results are highly insignificant. This is because students from poorer families tend to perform less well at KS2, so that there are very few FSM-eligible pupils at the top of the age-11 test score distribution. All in all, we find little evidence of remarkably heterogeneous peer effects along the dimension of family income.

7. 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 (9th grade) pupils and measuring peers' quality by their academic ability as recorded by test scores at age-11 (6th grade). In order to shed some light on the nature of peer effects, we have estimated both the effect of average peer quality, as well as the effect of being at school with a high proportion of very low-ability and very high-ability pupils, on the cognitive outcomes of regular students.

We view our main methodological contribution as twofold. Firstly, we 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 based on new peers in secondary schools we were 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 control for family and school × cohort unobservables, and pupil ability that is constant across subjects. As some degree of subject-specific sorting 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 have presented a set of robustness checks that support the causal interpretation of our findings. Additionally, we have proposed to focus on a 'limit 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 pseudo-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. Provided that the amount of sorting on subject-specific observables provides guidance on the amount of selection on unobservable subject-specific attributes, in particular lagged test scores, this sample should limit to the minimum the chances that our estimates are biased because of subject-specific endogenous sorting of pupils or omitted variables.

In terms of findings, 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 average peer quality effect 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 from these interactions. Finally, we find 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, a proxy for family income. This backs the intuition that our results are driven by peers' academic ability, and not predominantly related to their family background.

As a more general remark, our findings are highly relevant because of their strong external validity. Our 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). Additionally, our sample is 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 relatively small and local randomized trials in terms of recovering heterogeneous effects and having strong external validity.

Finally, the peer effect estimates that we present here 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 one 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 similar to the effects of full tracking of pupils by ability based on the experimental evidence presented in Duflo et al. (2008), with the caveat that their findings come from multi-age classes in primary schools in Kenya.

Do our results lend *overall* support to tracking of students by ability? Besides any equity consideration, there is no simple answer to this question from an efficiency-of-learning 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 in particular in relation according to one pupils' ability and gender. For example, pupils in the bottom of the ability distribution, and in particular girls, would not necessarily benefit (nor loose out) from more homogenous schools, as the negative peer effect of the bottom 5% students is counterbalanced by the beneficial effect of the top 5% peers. The opposite is true for students at the top of the ability distribution, and for boys in general, who would significantly *gain* from not interacting with the very weak and very academically bright students. In conclusion, 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.

8. References

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

Altonji, G., T. Elder and C. Taber (2005), "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 113, 151-184.

Ammermueller, A and J. Pischke (2009): "Peer Effects in European Primary Schools: Evidence from PIRLS," *Journal of Labor Economics*, 27, 315-348.

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: "The Effect of High-Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, September 2009.

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

Bandiera, O. , I. Barankay, I. Rasul (forthcoming): "Social Incentives in the Workplace," *Review of Economic Studies*.

Bandiera, O., V. Larcinese and I. Rasul (2008): "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students," mimeo, London School of Economics.

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.

De Giorgi, G., M. Pellizzari and S. Redaelli (2009): "Be As Careful of the Company that You Keep As of the Books You Read: Peer Effects in Education and on the Labor Market,", NBER Working Paper 14948.

DfES (2006): "Grouping Pupils for Success," *Primary and Secondary National Strategies*, Department for Education and Skills, London.

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.

Glennerster, H. (1991), “Quasi-Markets for Education,” *Economic Journal*, 101, 1268-76.

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.

Imbens, G. and T. Lemieux (2008): “Regression Discontinuity Designs: A Guide to Practice”, *Journal of Econometrics*, vol. 142(2), 615-635.

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.

Kutnick, P., J. Sebba, P. Blatchford, M. Galton, J. Thorp (2005): “The Effects of Pupil Grouping: Literature Review,” Department for Education and Skills Research Report 688.

Kutnick, P., S. Hodgekinson, J. Sebba, S. Humphreys, M. Galton, S. Steward, P. Bltachford and E. Baines (2006): “Pupil Grouping Strategies and Practices at Key Stage 2 and 3: Case Studies of 24 Schools in England,” Department for Education and Skills Research Report 796.

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.

West, A. and A. Hind (2003), "Secondary school admissions in England: Exploring the extent of overt and covert selection", Report for Research and Information on State Education Trust, www.risetrust.org.uk/admissions.html

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

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

Variable	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
Average peer achievement at KS2 in English, *new peers*	49.79	8.71	1	98
Average peer achievement at KS2 in Maths, *all peers*	50.16	6.85	23.94	72.82
Average peer achievement at KS2 in Math, *new peers*	49.94	8.06	1	100
Average peer achievement at KS2 in Science, *all peers*	49.93	7.35	25.31	73.83
Average peer achievement at KS2 in Science, *new peers*	49.68	8.35	1	100
Panel B: Top 5% treatments				
Fraction of top 5% in English, *all peers*	5.14	3.19	0	19.56
Fraction of top 5% in English, *new peers*	4.22	3.03	0	19.56
Fraction of top 5% in Maths, *all peers*	4.63	2.70	0	19.87
Fraction of top 5% in Maths, *new peers*	3.77	2.60	0	19.87
Fraction of top 5% in Science, *all peers*	4.84	2.92	0	19.86
Fraction of top 5% in Science, *new peers*	3.91	2.75	0	19.86
Panel C: Bottom 5% treatments				
Fraction of bottom 5% in English, *all peers*	4.64	3.00	0	19.30
Fraction of bottom 5% in English, *new peers*	3.79	2.78	0	19.30
Fraction of bottom 5% in Maths, *all peers*	4.68	2.86	0	19.86
Fraction of bottom 5% in Maths, *new peers*	3.81	2.67	0	19.86
Fraction of bottom 5% in Science, *all peers*	4.59	3.10	0	19.78
Fraction of bottom 5% in Science, *new peers*	3.78	2.90	0	19.78
Percentages of new peers for pupils in Year 7	87.56	22.66	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 – Impact of peer quality on KS3 educational attainments: main results

	Average achievement at KS2 (percentiles)		*Percentage of bottom 5% pupils*		*Percentage of top 5% pupils,*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
KS3 percentiles, unconditional on KS2	0.366	0.094	-0.959	-0.214	0.785	0.147
	(0.012)**	(0.014)**	(0.029)**	(0.036)**	(0.028)**	(0.029)**
KS3 percentiles, controlling for KS2 same-subject	0.305	0.058	-0.680	-0.118	0.401	0.061
	(0.010)**	(0.013)**	(0.026)**	(0.034)**	(0.024)**	(0.027)*
KS3 percentiles, controlling for KS2 same- and cross-subject	0.284	0.058	-0.615	-0.118	0.376	0.061
	(0.010)**	(0.013)**	(0.025)**	(0.034)**	(0.024)**	(0.027)*
KS3 percentiles, controlling for KS2 same- and cross-subject,	0.279	0.058	-0.611	-0.124	0.374	0.064
interacted with subject dummies	(0.010)**	(0.013)**	(0.025)**	(0.034)**	(0.024)**	(0.027)**
KS3 percentiles, controlling for KS2 same- and cross-subject,	0.264	0.056	-0.581	-0.112	0.384	0.048
interacted with subject dummies, and KS1 test scores	(0.010)**	(0.014)**	(0.026)**	(0.038)**	(0.025)**	(0.030)

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. All regressions control for quality of old peers. Pupil characteristics controlled for in Columns 1, 3 and 5; absorbed in Columns 2, 4 and 6. N. of observations: approx. 3,838,000 (1,279,000 pupils). N. of schools: 2194. KS1 data only available for 3 cohorts. KS1 test score in Science not available: imputed using pupil's average between KS1 Mathematics and English. Standard error clustered at the school level. **: at least 1%; *: at least 5%.

Table 4 – Impact of peer quality on KS3 educational attainments: robustness to school selectivity and potential pupils' subject-specific sorting

	Average achievement at KS2 (percentiles)		*Percentage of bottom 5% pupils*		*Percentage of top 5% pupils*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-estimator	OLS	Within-estimator	OLS	Within-estimator
KS3 percentiles, controlling for KS2: excluding specialist schools	0.283	0.061	-0.620	-0.131	0.377	0.073
	(0.010)**	(0.014)**	(0.027)**	(0.036)**	(0.025)**	(0.028)**
KS3 percentiles, controlling for KS2: oversubscribed schools (excluding specialist)	0.245	0.052	-0.633	-0.121	0.290	0.076
	(0.018)**	(0.023)*	(0.047)**	(0.069)*	(0.039)**	(0.041)*
KS3 percentiles, controlling for KS2: undersubscribed schools (excluding specialist)	0.284	0.067	-0.573	-0.133	0.421	0.076
	(0.013)**	(0.018)**	(0.032)**	(0.041)**	(0.033)**	(0.038)*
KS3 percentiles, controlling for KS2: sample of pupils whose *best* subject is the	0.287	0.051	-0.584	-0.102	0.378	0.058
same as best subjects of new peers in school on average (sorted)	(0.011)**	(0.013)**	(0.028)**	(0.036)**	(0.025)**	(0.028)**
KS3 percentiles, controlling for KS2: sample of pupils whose *best* subject is	0.274	0.054	-0.621	-0.140	0.380	0.071
different from best subject of new peers in school on average (mixed)	(0.010)**	(0.014)**	(0.025)**	(0.035)**	(0.025)**	(0.027)**
KS3 percentiles, controlling for KS2: replacing pupil fixed effects with secondary school × subject and secondary school × year fixed effects	0.279	0.141	-0.611	-0.114	0.374	0.121
	(0.010)**	(0.004)**	(0.025)**	(0.010)**	(0.024)**	(0.007)**
KS3 percentiles, controlling for KS2: including pupils fixed effects and school × subject fixed effects	0.286	0.014	-0.615	-0.072	0.386	0.033
	(0.012)**	(0.005)**	(0.029)**	(0.013)**	(0.029)**	(0.010)**

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. Estimates of the effect of average peer achievement at KS2 obtained from a separate regression. The table displays coefficients on treatments based on new peers. All regressions control for quality of old peers. Controls for KS2 in same- and cross-subject in interaction with subject dummies included in all regressions. Pupil characteristics controlled for throughout (absorbed with pupil fixed effects). Specialist schools account for about 8.5% of the pupil sample. Oversubscribed schools enrol approximately 40% of pupils in non-specialist schools. Sample of pupils with same best subject as new peers in school on average account for about 34% of the full sample. Sample of pupils with different best subject from new peers in school on average account for about 59% of the full sample. Remaining pupils do not have a clear subject ranking. Regression including school × subject fixed effect only considers the first cohort (year 7 in 2002) and last cohort (year 7 in 2005). Standard error clustered at the school level, except last two rows, Columns (2), (4) and (6), where they are robust. **: at least 1% significant; *: at least 5% significant.

Table 5 – Balancing of individual characteristics with respect to treatments; restricted sample with std.dev.≤3

	Average achievement at KS2 (percentiles)	*Percentage of top 5% pupils*	*Percentage of bottom 5% pupils*
Dependent variable is:	(1)	(3)	(5)
Panel A: OLS regression results			
First language is English	0.003	-0.014	0.000
	(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	0.292	-1.152	1.101
	(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)

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. **: at least 1% significant; *: at least 5% 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 6 – Impact of peer quality on KS3 educational attainments: main results; restricted sample with std.dev.≤3

	Average achievement at KS2 (percentiles)		*Percentage of bottom 5% pupils*		*Percentage of top 5% pupils,*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
KS3 percentiles, unconditional on KS2	0.405	0.046	-1.223	-0.148	1.092	0.070
	(0.018)**	(0.016)**	(0.045)**	(0.040)**	(0.047)**	(0.033)*
KS3 percentiles, controlling for KS2 same-subject	0.224	0.045	-0.486	-0.145	0.323	0.068
	(0.010)**	(0.016)**	(0.026)**	(0.040)**	(0.026)**	(0.033)*
KS3 percentiles, controlling for KS2 same- and cross-subject	0.223	0.045	-0.484	-0.145	0.321	0.068
	(0.010)**	(0.016)**	(0.026)**	(0.040)**	(0.026)**	(0.033)*
KS3 percentiles, controlling for KS2 same- and cross-subject,	0.224	0.055	-0.482	-0.149	0.330	0.097
interacted with subject dummies	(0.010)**	(0.016)**	(0.026)**	(0.040)**	(0.026)**	(0.033)**
KS3 percentiles, controlling for KS2 same- and cross-subject,	0.213	0.047	-0.457	-0.117	0.350	0.096
interacted with subject dummies and KS1 test scores	(0.011)**	(0.018)**	(0.028)**	(0.047)*	(0.028)**	(0.039)*

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. All regressions control for quality of old peers. Pupil characteristics controlled for in Columns 1, 3 and 5; absorbed in Columns 2, 4 and 6. Standard error clustered at the school. **: at least 1% significant; *: at least 5% significant. N. of observations: approx. 230,000 (77,000 pupils). N. of schools: 2194. KS1 data only available for 3 cohorts. KS1 test score in Science not available: imputed using pupil's average between KS1 Mathematics and English.

Table 7 – Impact of peer quality on KS3 educational attainments: top 5%, bottom 5% and average new peer quality included simultaneously; selected specifications

	Average peer KS2	*Bottom 5% pupils*	*Top 5% pupils*
Dependent variable is:	(1)	(2)	(3)
Panel A: Average peer quality includes top/bottom 5% pupils			
KS3 percentiles, controlling for KS2 same-subject	0.038	-0.124	0.058
	(0.013)**	(0.035)**	(0.028)*
KS3 percentiles, controlling for KS2 same- and cross-subject	0.038	-0.124	0.058
	(0.013)**	(0.035)**	(0.027)*
KS3 percentiles, controlling for KS2 same- and cross-subject, interacted with subject dummies	0.037	-0.128	0.061
	(0.014)**	(0.034)**	(0.027)*
Panel B: Average peer quality excludes top/bottom 5% pupils			
KS3 percentiles, controlling for KS2 same-subject	0.039	-0.129	0.061
	(0.012)**	(0.035)**	(0.027)*
KS3 percentiles, controlling for KS2 same- and cross-subject	0.039	-0.129	0.061
	(0.012)**	(0.035)**	(0.027)*
KS3 percentiles, controlling for KS2 same- and cross-subject, interacted with subject dummies	0.037	-0.133	0.064
	(0.012)**	(0.034)**	(0.027)*

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. Estimates come from one single regression including simultaneously all new peers quality measures and controlling control for old peers quality. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant. Number of observations: approximately 3,800,000 (1,260,000 pupils). Number of schools: 2194.

Table 8 – Impact of peer quality on KS3 attainments: treatments separately defined by pupils' free school meal eligibility

	Percentage of bottom 5% pupils		*Percentage of top 5% pupils*	
	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)
KS3 percentiles,	-0.242	-0.071	0.202	0.057
controlling for KS2	(0.065)**	(0.042)§	(0.141)	(0.027)*
Effect size	*0.999*	*0.467*	*0.428*	*0.627*

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 computed separately for pupils eligible for free school meals or not. Controls for KS2 in same- and cross-subject in interaction with subject dummies included in all regressions. Effect size (in *italics*) refer to the effect of a one standard deviation of the within-pupil distribution of peers as a percentage of one standard deviation of the within-pupil distribution of KS3 percentiles. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; §: at least 10% significant. Number of observations: approximately 3,800,000 (1,260,000 pupils). Number of schools: 2194.

Table 9 – Impact of peer quality on KS3 attainments: by pupil's ability

	Full sample		
	Average peer KS2	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
Dependent variable is: KS3, controlling for KS2	(1)	(2)	(3)
Effect for percentile 5-20	0.036	-0.119	0.059
	(0.013)**	(0.030)**	(0.027)*
Effect for percentile 20-35	0.034	-0.101	0.089
	(0.015)*	(0.037)**	(0.031)**
Effect for percentile 35-50	0.036	-0.112	0.075
	(0.016)*	(0.042)**	(0.034)*
Effect for percentile 50-65	0.038	-0.160	0.072
	(0.016)*	(0.044)**	(0.034)*
Effect for percentile 65-80	0.033	-0.152	0.032
	(0.016)*	(0.044)**	(0.033)
Effect for percentile 80-95	0.050	-0.080	0.009
	(0.017)**	(0.052)	(0.033)
F-Test: all coeffs. jointly equal to zero (p-value)	*0.0498*	*0.0009*	*0.0863*
F-Test: all coefficients are equal (p-value)	*0.9320*	*0.2216*	*0.1949*

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 peer quality are obtained from one single regression including all treatments. The table displays the coefficient on treatments based on new peers. All regressions control for the quality of old peers and interactions with ability groups. Controls for KS2 in same- and cross-subject in interaction with subject dummies are included in all regressions. Number of observations: 3,618,702 (1,206,234) pupils; corresponding to 2194 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; §: at least 10% significant. Interaction terms obtained by interacting the peer quality measures (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*. Note that the effect of KS2 achievement (same- and cross-subject) is controlled for semi-parametrically by interacting pupil KS2 percentiles with the dummies indicating his/her rank in the ability distribution (and in interaction with subject dummies). Ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution.

Table 10 – Impact of peer quality on KS3 attainments, by pupil's ability and gender

	Boys only			Girls only		
	Average peer KS2	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*	*Average peer KS2*	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
Dependent variable is: KS3, controlling for KS2	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pupils of ability pooled (overall effect)						
Overall effect	-0.007	-0.076	-0.052	0.026	-0.098	0.066
	(0.013)	(0.035)*	(0.028)§	(0.014)§	(0.037)**	(0.029)*
Panel B: Ability blocks defined on original KS2 percentiles						
Effect for	-0.009	-0.093	-0.013	0.043	-0.080	0.066
percentile 5-20	(0.013)	(0.032)**	(0.029)	(0.016)**	(0.038)*	(0.035)§
Effect for	-0.006	-0.057	-0.037	0.025	-0.072	0.126
percentile 20-35	(0.016)	(0.039)	(0.033)	(0.018)	(0.044)§	(0.037)**
Effect for	-0.012	-0.068	-0.059	0.027	-0.066	0.088
percentile 35-50	(0.017)	(0.046)	(0.036)	(0.018)	(0.047)	(0.039)*
Effect for	-0.010	-0.106	-0.036	0.026	-0.113	0.062
percentile 50-65	(0.018)	(0.048)*	(0.038)	(0.018)	(0.050)*	(0.038)§
Effect for	-0.009	-0.089	-0.079	0.019	-0.139	0.023
percentile 65-80	(0.018)	(0.051)§	(0.037)*	(0.018)	(0.050)**	(0.036)
Effect for	0.030	0.036	-0.096	0.016	-0.116	0.011
percentile 80-95	(0.022)	(0.065)	(0.043)*	(0.020)	(0.060)*	(0.039)
F-Test: all coeff. jointly equal to zero (p-value)	*0.6604*	*0.0425*	*0.2642*	*0.2765*	*0.1042*	*0.0334*
F-Test: all coefficients are equal (p-value)	*0.5531*	*0.2597*	*0.4281*	*0.8037*	*0.6809*	*0.0766*

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the peer quality effects obtained from one single regression including all treatments. The table displays the coefficient on treatment based on new peers. All regressions control for the quality of old peers (and interactions with ability groups in Panel B). Controls for KS2 in same- and cross-subject in interaction with subject dummies are included in all regressions. Two separate regressions run for boys and girls. Number of observations for boys: 1,814,310 (604,770 pupils) in 2101 schools. Number of observations for girls: 1,804,392 (601464 pupils) in 2134 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; §: at least 10% significant. Interaction terms in Panel B obtained by interacting the peer quality measures (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*. Note that the effect of KS2 achievement (same- and cross-subject) is controlled for semi-parametrically by interacting pupil KS2 percentiles with the dummies indicating his/her rank in the ability distribution (and in interaction with subject dummies). Ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution.

Table 11 – Impact of peer quality on KS3 attainments, by pupil's ability and free school meal eligibility

	Pupil is eligible for free school meals			Pupil is not eligible for free school meals		
	Average peer KS2	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*	*Average peer KS2*	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
Dependent variable is: KS3, controlling for KS2	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Pupils of ability pooled (overall effect)						
Overall effect	0.056	-0.155	0.067	0.034	-0.114	0.061
	(0.022)*	(0.044)**	(0.042)	(0.013)**	(0.036)**	(0.028)*
Panel B: Ability blocks defined on original KS2 percentiles						
Effect for	0.041	-0.143	0.076	0.033	-0.102	0.055
percentile 5-20	(0.021)*	(0.041)**	(0.042)§	(0.013)**	(0.033)**	(0.029)§
Effect for	0.050	-0.144	0.064	0.031	-0.078	0.095
percentile 20-35	(0.028)§	(0.052)**	(0.052)	(0.016)*	(0.038)*	(0.032)**
Effect for	0.063	-0.116	0.092	0.032	-0.102	0.074
percentile 35-50	(0.031)*	(0.064)§	(0.064)	(0.016)*	(0.044)*	(0.034)*
Effect for	0.052	-0.201	0.113	0.036	-0.146	0.068
percentile 50-65	(0.038)	(0.074)*	(0.068)§	(0.016)*	(0.046)**	(0.035)*
Effect for	0.073	-0.196	-0.029	0.030	-0.138	0.037
percentile 65-80	(0.043)§	(0.089)*	(0.079)	(0.016)§	(0.046)**	(0.033)
Effect for	0.103	-0.203	-0.041	0.047	-0.061	0.011
percentile 80-95	(0.068)	(0.141)	(0.126)	(0.017)**	(0.053)	(0.034)
F-Test: all coeff. jointly equal to zero (p-value)	*0.2857*	*0.0093*	*0.3031*	*0.1035*	*0.0112*	*0.1075*
F-Test: all coefficients are equal (p-value)	*0.9128*	*0.8929*	*0.5834*	*0.9275*	*0.2050*	*0.2271*

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variables on treatments. Estimates of the peer quality effects obtained from one single regression including all treatments. The table displays the coefficient on treatment based on new peers. All regressions control for the quality of old peers (and interactions with ability groups in Panel B). Controls for KS2 in same- and cross-subject in interaction with subject dummies are included in all regressions. Two separate regressions were run for pupils eligible and non-eligible for free school meals. Number of observations regressions including pupils eligible for free school meals only: 468,009 (156,003 pupils) in 2193 schools. Number of observations regressions including pupils non-eligible for free school meals only: 3,150,693 (1,050,231 pupils) in 2193 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; §: at least 10% significant. Interaction terms in Panel B obtained by interacting the peer quality measures (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*. Note that the effect of KS2 achievement (same- and cross-subject) is controlled for semi-parametrically by interacting pupil KS2 percentiles with the dummies indicating his/her rank in the ability distribution (and in interaction with subject dummies). Ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution.

Appendix Tables

Appendix Table 1 – Within and between variation in pupil test scores and treatment measures

	Full Sample				Restricted sample with std.dev.≤3				Restricted sample with std.dev.≤11.5			
Variable:	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*
KS2 percentiles	49.19	24.31	21.15	11.98	41.95	29.88	29.83	1.68	47.09	25.80	25.10	5.97
KS3 percentiles	49.10	25.61	22.99	11.29	42.98	29.72	28.36	8.88	47.59	27.11	25.21	9.97
Average peer achievement at KS2	49.80	8.38	7.96	2.61	49.46	8.39	7.98	2.58	49.74	8.39	7.98	2.60
Fraction of bottom 5%	3.79	2.78	2.62	0.94	3.90	2.84	2.67	0.95	3.90	2.84	2.67	0.95
Fraction of top 5%	3.97	2.81	2.49	1.29	3.90	2.78	2.47	1.28	3.96	2.80	2.49	1.29

Note: Number of observations in full sample: 3,838,542 corresponding to 1,279,514 pupils and 3 subjects. Number of observations in restricted sample with std.dev.≤3: 231,966 corresponding to 77,322 pupils and 3 subjects. Number of observations in restricted sample with std.dev.≤11.5: 1,866,516 corresponding to 622,172 pupils and 3 subjects. Peer quality measures refer to new peers only.

Appendix Table 2 – 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 by 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 3 – Additional descriptive statistics: full sample and restricted samples with std.dev. ≤3 and std.dev.≤ 11.5

Variable	Full Sample	Restricted sample (std.dev.≤3)	Restricted Sample (std.dev.≤11.5)
Panel A: Pupils' Outcomes			
KS2 percentile, English	49.3 (24.3)	42.0 (29.8)	47.3 (25.6)
KS2 percentile, Mathematics	49.4 (24.3)	42.0 (29.9)	47.2 (25.8)
KS2 percentile, Science	48.9 (24.3)	41.9 (29.9)	46.8 (26.0)
KS3 percentile, English	48.9 (26.0)	43.0 (29.4)	47.6 (27.2)
KS3 percentile, Mathematics	49.2 (25.3)	42.6 (29.8)	47.4 (26.9)
KS3 percentile, Science	49.2 (25.5)	43.2 (29.9)	47.7 (27.3)
Panel B: Pupils' Characteristics			
First language is English	0.93 (0.25)	0.93 (0.26)	0.93 (0.25)
Eligible for free school meals	0.13 (0.34)	0.16 (0.36)	0.14 (0.34)
Male	0.50 (0.50)	0.49 (0.50)	0.49 (0.50)
Changed school between Year 7 and KS3	0.11 (0.31)	0.11 (0.32)	0.11 (0.31)
Ethnicity: White British	0.85 (0.35)	0.85 (0.36)	0.85 (0.35)
Ethnicity: White other	0.02 (0.12)	0.02 (0.13)	0.02 (0.12)
Ethnicity: Asian	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)
Ethnicity: Black	0.03 (0.16)	0.03 (0.16)	0.03 (0.16)
Ethnicity: Chinese	0.00 (0.05)	0.00 (0.04)	0.00 (0.04)
Ethnicity: Other	0.05 (0.22)	0.05 (0.22)	0.05 (0.22)
Panel C: School characteristics (Year 7)			
Cohort size	201.7 (57.2)	201.9 (57.4)	201.9 (57.2)
Community school	0.67 (0.47)	0.68 (0.47)	0.67 (0.47)
Voluntary aided school	0.14 (0.352)	0.14 (0.34)	0.14 (0.35)
Voluntary controlled school	0.03 (0.18)	0.03 (0.18)	0.03 (0.18)
Foundation school	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)
City Technology college school	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
Religiously affiliated school	0.16 (0.37)	0.15 (0.36)	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. Restricted samples are composed of pupils with standard deviation of KS2 percentiles across subjects ≤3 and ≤11.5. Number of pupils in restricted sample std.dev.≤3: 77,322. Number of pupils in restricted sample with std.dev.≤11.5: 622,172. Full sample and restricted samples 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.

Appendix Table 4 – Impact of peer quality on KS3 educational attainments; restricted sample with std.dev.≤ 11.5

	Average achievement at KS2 (percentiles)		*Percentage of bottom 5% pupils*		*Percentage of top 5% pupils,*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
KS3 percentiles, unconditional on KS2	0.387	0.070	-1.061	-0.152	0.898	0.089
	(0.013)**	(0.014)**	(0.032)**	(0.034)**	(0.031)**	(0.027)**
KS3 percentiles, controlling for KS2 same-subject	0.267	0.060	-0.582	-0.121	0.363	0.065
	(0.010)**	(0.013)**	(0.024)**	(0.033)**	(0.023)**	(0.026)*
KS3 percentiles, controlling for KS2 same- and cross-subject	0.260	0.060	-0.563	-0.121	0.355	0.065
	(0.009)**	(0.013)**	(0.024)**	(0.033)**	(0.023)**	(0.026)*
KS3 percentiles, controlling for KS2 same- and cross-subject,	0.260	0.066	-0.562	-0.126	0.358	0.082
interacted with subject dummies	(0.009)**	(0.013)**	(0.024)**	(0.033)**	(0.023)**	(0.026)**
KS3 percentiles, controlling for KS2 same- and cross-subject,	0.247	0.063	-0.535	-0.118	0.368	0.064
interacted with subject dummies, and KS1 test scores	(0.010)**	(0.014)**	(0.025)**	(0.037)**	(0.024)**	(0.030)*

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. All regressions control for quality of old peers. Pupil characteristics controlled for in Columns 1, 3 and 5; absorbed in Columns 2, 4 and 6. Standard error clustered at the school. **: at least 1% significant; *: at least 5% significant. N. of observations: approx. 1,860,000 (620,000 pupils). N. of schools: 2194. KS1 data only available for 3 cohorts. KS1 test score in Science not available: imputed using pupil's average between KS1 Mathematics and English. N. of observations when controlling for KS1: approx. 1,290,000 (430,000 pupils). N. of schools: 2194.

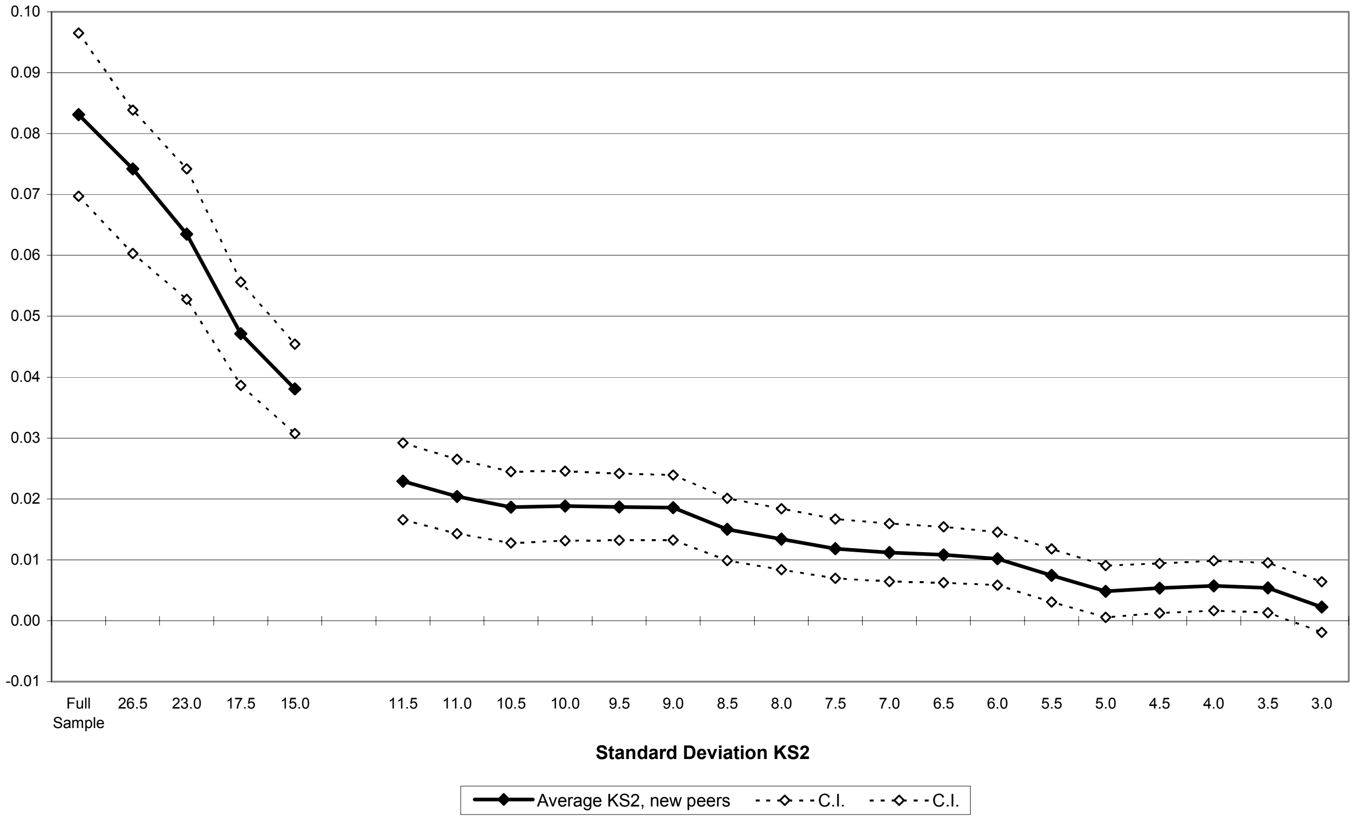
Appendix Table 5 – Impact of peer quality on KS3 educational attainments: by pairs of subjects; full sample

	Average achievement at KS2 (percentiles)		*Percentage of bottom 5% pupils*		*Percentage of top 5% pupils,*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
KS3 percentiles, controlling for KS2:	0.142	0.057	-0.498	-0.151	0.235	0.125
English and Maths only	(0.010)**	(0.021)**	(0.031)**	(0.062)**	(0.028)**	(0.045)**
KS3 percentiles, controlling for KS2:	0.153	0.045	-0.511	-0.175	0.225	0.032
English and Science only	(0.012)**	(0.022)*	(0.031)**	(0.060)**	(0.025)**	(0.048)
KS3 percentiles, controlling for KS2:	0.145	-0.003	-0.525	-0.055	0.255	0.011
Maths and Science only	(0.011)**	(0.016)	(0.027)**	(0.040)	(0.027)**	(0.031)

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 average peer achievement at KS2 and of the percentage of top 5% pupils and percentage of bottom 5% pupils are obtained from one single regression including the three treatments simultaneously. The table displays the coefficients on treatments based on new peers. All regressions control for quality of old peers. Controls for KS2 in same subject in interaction with subject dummies and controls for KS2 in cross-subjects included in all regressions. Pupil characteristics controlled as in Table 4. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant. N. of observations: approx. 2,410,000 (1,205,000). N. of schools 2194.

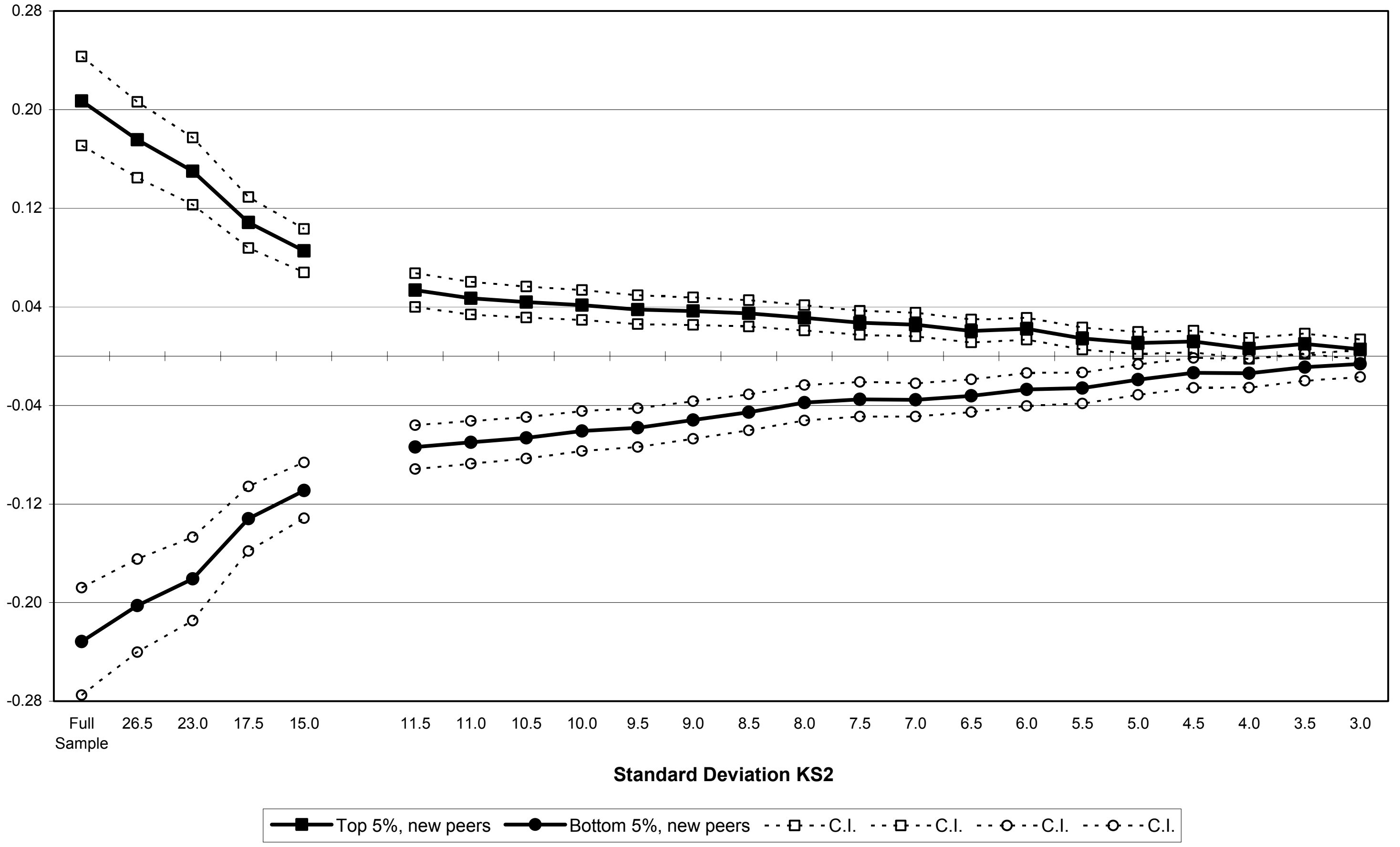
Figures

Figure 1 – Balancing of KS2 with respect to treatment, by cumulative bands of standard deviation of KS2attainments (percentiles)



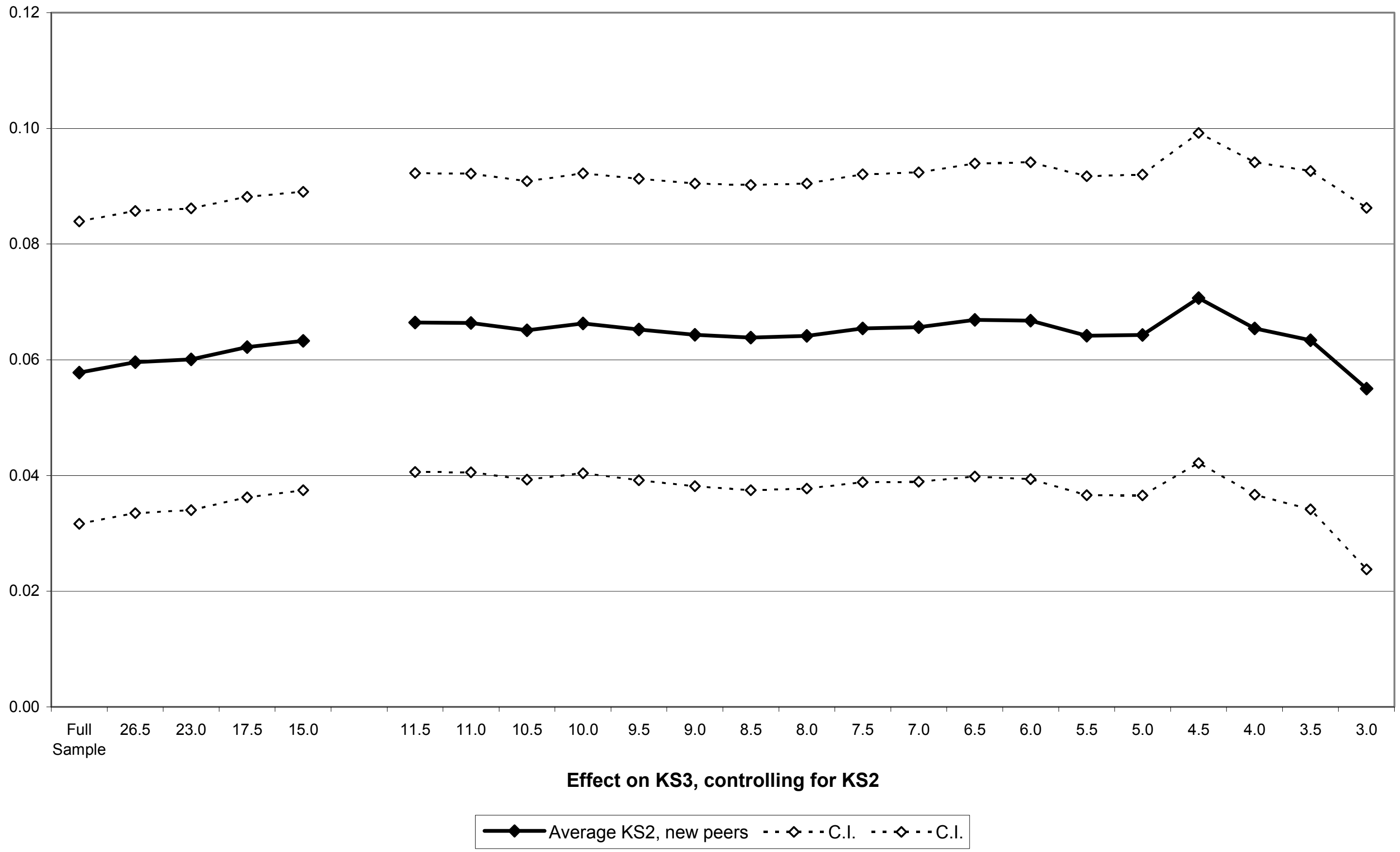
Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS2 achievements (percentiles) on the average achievement at KS2 of new peers. All regressions include pupil fixed effects and control for old peer quality. 23 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, and then std.dev.≤15; std.dev.≤17.5; std.dev.≤23; std.dev.≤26; full sample. 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; std.dev.≤15 includes roughly 62.5% of the full sample; std.dev.≤17 includes roughly 75% of the full sample; std.dev.≤23 includes roughly 90% of the full sample; std.dev.≤26.5 includes roughly 95% of the full sample.

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



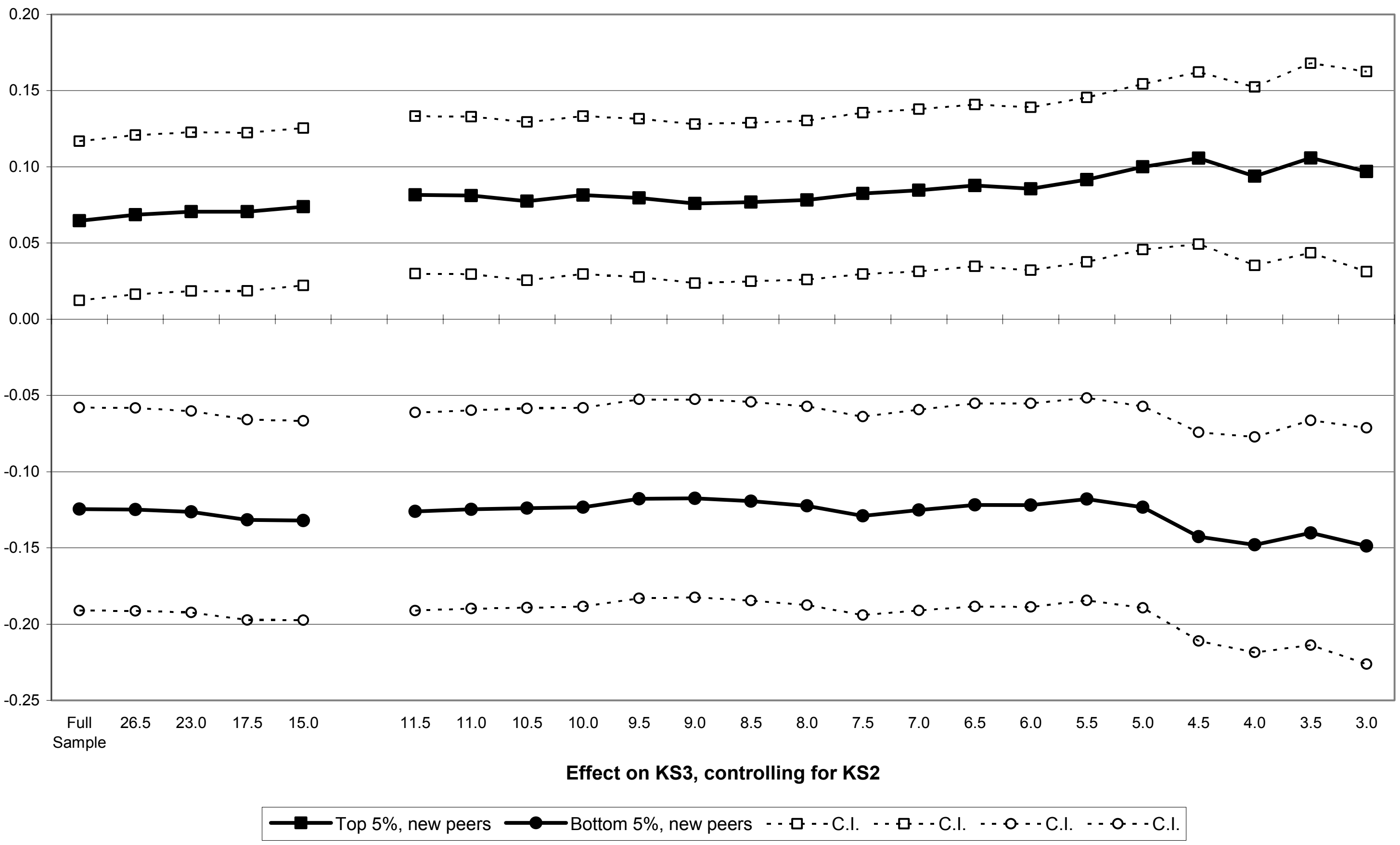
Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS2 achievements (percentiles) on the percentage of top 5% pupils, new peers, and percentage of bottom 5% pupils, new peers. All regressions include pupil fixed effects and control for old peer quality. 23 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, and then std.dev.≤15; std.dev.≤17.5; std.dev.≤23; std.dev.≤26; full sample. 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; std.dev.≤15 includes roughly 62.5% of the full sample; std.dev.≤17 includes roughly 75% of the full sample; std.dev.≤23 includes roughly 90% of the full sample; std.dev.≤26.5 includes roughly 95% of the full sample.

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



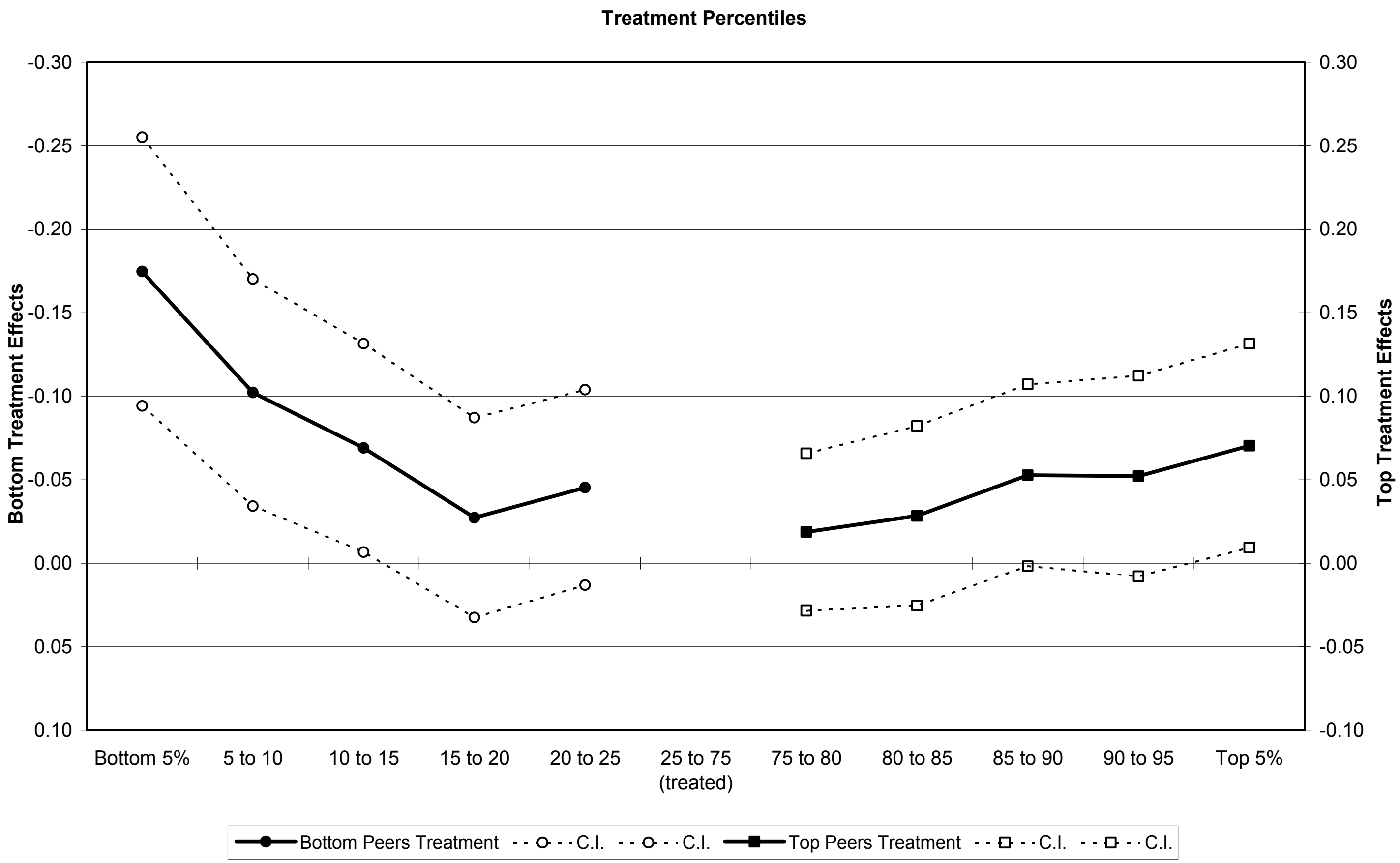
Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS3 achievements (percentiles) on the average achievement at KS2 of new peers. All regressions include pupil fixed effects, control for old peer quality and pupil KS2 achievement (percentiles). 23 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, and then std.dev.≤15; std.dev.≤17.5; std.dev.≤23; std.dev.≤26; full sample. 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; std.dev.≤15 includes roughly 62.5% of the full sample; std.dev.≤17 includes roughly 75% of the full sample; std.dev.≤23 includes roughly 90% of the full sample; std.dev.≤26.5 includes roughly 95% 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). 23 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, and then std.dev.≤15; std.dev.≤17.5; std.dev.≤23; std.dev.≤26; full sample. 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; std.dev.≤15 includes roughly 62.5% of the full sample; std.dev.≤17 includes roughly 75% of the full sample; std.dev.≤23 includes roughly 90% of the full sample; std.dev.≤26.5 includes roughly 95% of the full sample.

Figure 5 – Effect of treatments on KS3 percentiles, conditional of KS2, by different percentile cut-off points for top and bottom peers



Note: The figure plots regression coefficients and confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS3 achievements on the following treatments: percentage of top 5% new peers; percentage of top 5-to-10% new peers; percentage of top 10-to15% new peers; percentage of top 15-to-20% new peers; percentage of top 20-to25% new peers; percentage of bottom 5% new peers; percentage of bottom 5-to-10% new peers; percentage of bottom 10-to15% new peers; percentage of bottom 15-to-20% new peers; percentage of bottom 20-to25% new peers. All regressions include pupil fixed effects, controls for old peer quality and pupil KS2 achievement in interaction with subject dummies and cross-subject KS2 achievements. Treated pupils include students with KS2 achievements between 25^{th} and 75^{th} percentile of the cohort-specific distribution of KS2 for every subjects. Number of observations: 2,726,310 (908,770 pupils) in 2194 schools.