

NBER WORKING PAPER SERIES

THE GOOD, THE BAD AND THE AVERAGE:
EVIDENCE ON THE SCALE AND NATURE OF ABILITY PEER EFFECTS IN SCHOOLS

Victor Lavy
Olmo Silva
Felix Weinhardt

Working Paper 15600
<http://www.nber.org/papers/w15600>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2009

We would like to thank Rebecca Allen, Josh Angrist, Kenneth Chay, Steve Gibbons, Francis Kramarz, Steve Machin, Michele Pellizzari, Steve Pischke, Steve Rivkin, Yona Rubinstein, Henry Overman, Hongliang (Henry) Zhang, and seminar participants at Bocconi University, Brown University, CEP-LSE, CMPO-Bristol University, CREST in Paris, EIEF in Rome, ESSLE 2009, Hebrew University of Jerusalem, IFS and IoE in London, Royal Holloway University, and Tel Aviv University for helpful comments and discussions. We also thank the Center for Economic Performance (CEP) at the London School of Economics (LSE) for seed money for this project and Felix Weinhardt acknowledges ESRC PhD funding. All remaining errors remain our own. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2009 by Victor Lavy, Olmo Silva, and Felix Weinhardt. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

NBER Working Paper No. 15600

December 2009

JEL No. I21,J18,J24

ABSTRACT

We study the scale and nature of ability peer effects in secondary schools in England. 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 census data for four cohorts of pupils taking 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 negative peer effects arising from students at the very bottom of the ability distribution, but little evidence that the average peer quality and the very top peers significantly affect pupils' academic achievements. However, these results mask some significant heterogeneity along the gender dimension, with girls significantly benefiting from the presence of very academically bright peers, and boys significantly losing out. We further provide evidence that the effect of the very best peers substantially varies by the ability of other pupils. On the other hand, the effect of the very worst peers is similarly negative and significant for boys and girls of all abilities.

Victor Lavy
Department of Economics
Hebrew University
Mount Scopus
91905 Israel
and Royal Holloway
University of London
and NBER
msvictor@mscc.huji.ac.il

Felix Weinhardt
Department of Geography and Environment
London School of Economics
Houghton Street
London WC2A 2AE
UK
f.j.weinhardt@lse.ac.uk

Olmo Silva
Department of Geography and Environment
London School of Economics
Houghton Street
London WC2A 2AE
UK
<http://personal.lse.ac.uk/silvao/>
o.silva@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 effects¹, 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. Our aims are both to investigate the size (i.e. the 'scale') of ability peer effects on the outcomes of secondary school students, and to explore which segments of the ability distribution of peers drive the impact of peer quality on pupils' achievements (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 – as opposed to the average peer quality drive any significant peer effect on the outcomes of other students.

To do so, we use data for all secondary schools in England for four cohorts of age-14 (9th 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 refecation 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

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 (2009a) 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.

analysis, we focus on the effect of new peers' ability on pupil achievement (controlling for old peers' quality), thus by-passing reflection problems.³

Our results show that a large fraction of 'bad' peers at school as identified by students in the bottom 5% of the ability distribution negatively and significantly affect the cognitive performance of other schoolmates. Importantly, we find that it is only the very bottom 5% students that (negatively) matter, and not 'bad' peers in other parts of the ability distribution. On the other hand, we uncover little evidence that the average peer quality and the share of very 'good' peers as identified by students in the top 5% of the ability distribution affect the educational outcomes of other pupils. However, these findings mask a significant degree of heterogeneity along the gender dimension. Indeed, we show that girls significantly benefit from interactions with very bright peers, and the more so if they are in the bottom half of the ability distribution. In marked contrast, boys are negatively affected by a larger proportion of academically outstanding peers at school, with this adverse effect being more evident for male students in the top part of the ability distribution. On the other hand, we find that the negative effect of the very weak students does not significantly vary by the ability of regular students, nor along the gender dimension. Finally, the effect of the average peer quality on pupil cognitive achievement is estimated to be zero for boys and girls, and for students of different abilities.

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) literature⁴ 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 sorting/selection might be taking place, with pupils and schools being affected by and/or responding to cohort-specific unobserved shocks to students' and schools' quality. Identification

³ Note that this does not imply that we are able to separate endogenous from exogenous peer effects (see Mankiw, 1993). We see this as a further and separate issue from reflection problems that arise from previous or simultaneous interactions among students that affect measures of peers' ability (see Sacerdote, 2001).

⁴ A number of recent studies have 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 (2009b).

⁵ A similar result is documented by Gibbons and Telhaj (2008) and Black et al. (2009).

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 (i.e. specifications including pupil fixed-effects) 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. We then study whether subject-to-subject variation in outcomes for the same student is systematically associated with the subject-to-subject variation in peers' ability.

One significant advantage of this approach is that by including pupil fixed-effects we are able to control for pupil own unobservable average ability across the three subjects, as well as for unmeasured family background characteristics. Additionally, we can partial out in a non-parametric way school-by-cohort fixed-effects and other more general cohort-specific unobserved shocks that might affect pupils' outcomes and peers' quality similarly across the three subjects. This seems particularly important given the evidence of year-on-year secondary school sorting highlighted here above. On the other hand, one potential threat to our identification strategy is the possibility that sorting occurs along the lines of subject-specific abilities, so that within-student across-subject variation in ability is correlated with the variation in peers' ability across subjects. However, as we shall see below, conditional on pupil fixed-effects, our results are virtually identical irrespective of whether or not we control for pupils' age-11 test scores, a proxy for students' subject-specific prior academic ability. This is because there is neither a sizeable nor a significant correlation between the within-student across-subject variation in age-11 achievements, and the variation in peers' ability across subjects. This suggests that specifications that include pupil fixed-effects effectively take care of most of the sorting of pupils and their peers into secondary education, and provide reliable causal estimates of ability peer effects. To further support this claim, we provide an extensive battery of robustness checks and falsification exercises that lend additional credibility to the causal interpretation of our results.

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, Gould et al., 2009b), within-school random variation (Hoxby, 2000; Hanushek et al., 2003; Ammermualler and Pischke, 2009, Gould et al, 2009a), instrumental variables (Goux and Maurin, 2007) or sub-group re-assignments (Katz et al., 2001 and Sanbonmatsu et al., 2006).⁷ Only recently, Lavy and Schlosser (2007), Lavy et al. (2008) and Duflo et al. (2008) have tried to enter the 'black box' of ability peer effects in Israel and Kenya, respectively, and have explicitly focused on understanding the mechanisms through which interactions could exert their effects. Duflo et al. (2008) exploits random assignment of pupils in primary schools in Kenya to classes by ability in order to identify peer effects. 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. 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 pupils in English secondary schools. The authors attempt to control for the endogenous sorting of pupils to secondary schools by allowing for primary and secondary school fixed-effect interactions and trends. However, this approach does not fully eliminate the correlation

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), Bifulco et al. (2008), Burke and Sass (2008), Carrell and Hoekstra (2008), Figlio (2007), Lefgren (2004), Nechyba and Vidgor (2007) and Vidgor and Nechyba (2004).

between pupils' own ability and peer quality, and their results provide little evidence of sizeable and significant peer effects.

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 As already mentioned, this allows to control for pupil unobservable average ability, unmeasured family background characteristics, school-by-cohort fixed effects and other more general cohort-specific shocks that are common to the three subjects. We believe this approach helps us to achieve a clean identification of the causal effect of peers' ability. 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. Ammermueller and Pischke, 2009; Hoxby, 2000; Gibbons and Telhaj, 2008; Gould et al., 2009a; 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 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 taking their age-14 national tests in the academic years 2003/2004-2006/2007, we estimate the following pupil fixed-effect equation:

⁸ Lavy (2009) uses the same approach to investigate the effect of instructional time on academic achievements, while Bandiera et al. (2009) use within-student across-subjects variation to study class size effects at university and Bandiera et al. (forthcoming) exploit within-worker over-time variation to analyse social incentives at work.

$$A_{iqst} = \alpha_i + \beta_q + \gamma_{st} + \beta_q \times Gender_i + \delta_1 P_{qst} + \delta_2 P_{qst}^h + \delta_3 P_{qst}^l + \varepsilon_{iqst} \quad (1)$$

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 \times cohort effect. We also include an interaction term between pupil's gender and subject specific effects which is meant to control for the well-documented gender disparities in achievements in different subjects (see Ellison and Swanson, 2009 and Fryer and Levitt, forthcoming), and the effect that these might have on pupils' and their peers' sorting into secondary schooling.⁹ Next, 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 (denoted as Key Stage 2, or KS2). On the other hand, P_{qs}^h and P_{qs}^l capture the fraction of very high-ability and the very low-ability peers in one students' cohort. More precisely, we choose the top and bottom 5% in the (cohort-specific) national distribution of KS2 test scores as the cut off points to determine P_{qs}^h and P_{qs}^l (more details in the data section). 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 coefficients of interest are δ_1 , which captures the effect of the average ability of peers on students' achievement; δ_2 , which measures the effect of the proportion of peers in the cohort who are in the top 5% of the national distribution of KS2 test scores; and δ_3 , which identifies the effect of the fraction of students who are in the bottom 5%. As discussed above, we are interested in the relative strength and significance of these three coefficients to determine which segments of the peer ability distribution drive any ability peer effect that we will document.

Note that one significant advantage of this approach is that pupil fixed-effects 'absorb' students' own unobservable average ability across subjects as well as unmeasured family background characteristics. Moreover, this specification allows to partial out in a fully non-parametric way school-by-cohort fixed effects (e.g. unobserved changes in school resources or head teacher), and other more general cohort-specific unobserved shocks (e.g. changes in the quality of primary schooling or in the quality of childcare facilities) that might affect pupils' outcomes and peers' quality similarly across the three subjects. This seems particularly important given the issues discussed in Arcidiacono et al.

⁹ We also tried specifications where we interact other pupil characteristics (e.g. eligibility for free school meals) with subject specific dummies, and found virtually identical results. However, we prefer the more parsimonious specification in Equation (1).

(2009) and given that, as highlighted in the Introduction, we find evidence of 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-by-year specific considerations might be taking place, or that cohort-specific shocks to pupil and school quality might have occurred.

Before moving on, three 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 the δ parameters 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. Third, results are unchanged when we use the absolute number of very weak and very good peers instead of their proportion.

3.2. *Dealing with potential threats to identification*

Although the strategy described so far allows us to effectively control for pupils' average ability across subjects, unobservable family background characteristics and school-by-cohort unobservable shocks, this setup does not preclude the possibility that selection and sorting of students in different schools is partly based on subject-specific ability and considerations. In particular, there might be some residual correlation between the within-student across-subject variation in age-11 prior achievements, capturing students' subject-specific abilities, and the variation in peers' quality across subjects.

Our main approach to account for such potential sorting is to control for pupils' KS2 test scores in all subjects in the within-pupil estimation. The underlying assumption is that the lagged test scores effectively capture any subject-specific abilities, 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:

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

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.¹⁰ Anticipating our findings below, we find that results from within-pupil specifications are virtually unaffected by whether or not we control for pupils' age-11 test scores. This is because there is neither a sizeable nor a significant correlation between the within-student across-subject variation in prior achievements, and the variation in peers' ability across subjects. Stated differently, conditional on pupil fixed-effects, peers' subject-specific quality measures are almost perfectly balanced with respect to pupils' own age-11 test scores, and specifications that include pupil fixed-effects effectively take care of the sorting of pupils and their peers into secondary education.

We further complement our core strategy with a set of robustness checks and alternative specifications that allow us to gauge the importance of subject-specific school selection and pupil sorting. For example, we include in some of our specifications school-by-subject fixed-effects to control for the sorting of pupils and their peers into schools based on subject-specific school unobservables. All these exercises provide strong support to the causal interpretation of our estimates.

3.3. *Measuring peers' ability*

A key requirement for our empirical approach is that the proxies of peer ability are based on 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 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 the 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 \times cohort \times 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 that our estimates are not sensitive to the inclusion of these variables.

¹⁰ 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.

Two final remarks are worth being made. First, we use information about the school that a pupil is attending at age-12 (7th 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 7th-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 peers’ ability measured at the grade and *not* at the class level because our data does not include class identifiers, and because class placement might be endogenous since school authorities may have some discretion in placing students in different classes within a grade. However, we do not see this as a restrictive compromise since the majority of schools do not group pupils with different subject-specific abilities in different classes at the early stages of secondary education (see more details in the next section). Therefore, the quality of peers within a grade is likely to be strongly correlated with the quality of peers within classes. On the other hand, if some degree of subject-specific streaming takes place so that our peer quality measures capture the peer quality actually experienced by pupils with some noise, our estimates will be downward biased and more properly interpreted as ‘intention-to-treat’ peer effects.¹¹

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 1st and 2nd 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 (6th grade), children leave the primary phase and go on to secondary school where they progress through KS3 (7th to 9th grade) and KS4 (10th to 12th 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,

¹¹ Note that our study does not suffer from measurement error due to incomplete information on pupils’ schoolmates as in Ammermueller and Pischke (2009).

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 attainments 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, two important issues are worth mentioning. First, the concept of 'class' is a rather hollow one in English secondary schools since students tend to be grouped with different pupils for different subjects. A second important aspect that characterizes English secondary education is the practice of 'ability setting', i.e. subject-specific streaming. Under these arrangements, secondary school pupils are initially taught in mixed-ability groups for an observation and acclimatization period of around a year, and then eventually 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 at some point between 7th grade and 9th grade, but only 53% from grade 7. These figures are much lower for English and Science respectively at: 46% (at some stage between 7th and 9th grade) and 34% (from 7th grade); and 59% (sometimes between 7th and 9th grade) and 44% (from 7th grade). In conclusion, two important features emerge from this brief discussion. First, because of the lack of clearly defined and stable classes during secondary education, students will predominantly interact with different peers in different subjects. Second, since ability setting is not strictly implemented, 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 education¹². 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 School Census (PLASC), which records information on pupil's gender, age, ethnicity, language skills, any special educational needs or disabilities, entitlement to free school meals and various other pieces of information, including the identity of the school attended during years other than those when pupils sit for their Key Stage 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).

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 from age-11 (6th grade) through to age-14 (9th grade), and to join this information to the PLASC data for every year of secondary schooling (7th to 9th 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 7th grade), and not after two years (in 9th 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. 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

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.

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 5th percentile or above the 95th percentile of the cohort-specific national distribution of KS2 test scores.

We have imposed a set of restrictions on our data in order to obtain a balanced panel of pupil information in a balanced panel of schools. First, 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 lose 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 in any subject 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 approximately 1,300,000 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 solely 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

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).

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 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 significant mean reversion in the achievements of very good and very bad peers between age-11 and age-14. To further substantiate this point, we have thoroughly analyzed the KS3 percentile ranking of pupils in the top 5% and bottom 5% of the KS2 achievement distribution. For all subjects, about 80% of the pupils ranking in the bottom 5% at KS2, still rank in the bottom 20% of the KS3 distribution, with approximately 70% of them concentrated in the bottom 10%. At the opposite extreme, around 80% of pupils ranking in the top 5% at KS2 remains in the top 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 reveal 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 more likely to be eligible for free school meals (a proxy for family income). The opposite is true for pupils with at least one 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 16% 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

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.

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 present some descriptive statistics of our ‘treatments’. Statistics are presented for new peers only. Note once again that 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 on 50 for all subjects. Panel B and Panel C, instead, present descriptive statistics for our proxies for ‘good’ and ‘bad’ new peers. By construction, the fractions of top 5% and bottom 5% ‘new peers’ in the incoming cohort are smaller than the corresponding fractions including all peers (at around 5% each in every subject). Note that all peer quality measures display quite a wide range of variation, although this mainly capture differences across schools. Nevertheless, Appendix Table 1 shows that the same pupil faces considerably different fractions of academically bright and weak students across different subjects, as well as a significant amount of within-pupil across subject dispersion in average peer’s age-11 test scores. This is the variation that our pupil fixed-effect regressions exploit to identify the effect of peer quality.

5. Results

5.1. *Effects of peers’ ability: main results*

We begin the discussion of our 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 in 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 four rows of the table come from a variety of specifications, which differ in the way they control for lagged test scores. In the first two rows, we report estimates unconditional on age-11 achievements, while the third row presents estimates where we include pupils’ own KS2 attainment in the same subject in interaction with subject dummies. This allows pupils’ lagged outcomes to affect age-14 test scores differently in different subjects. Finally, in the last row, we include pupils’ own KS2 test scores in the *same*-subject and *cross*-subject (as detailed in Section 3.2) in interaction with subject effects to control for pupils’ own subject-specific ability, as well as his/her ‘spread’ of abilities across subjects and cross-subject spillovers. Note also that the results in the first row are obtained from different regressions entering either the average quality of peers, or the fraction of top 5% and bottom 5% peers in the grade. Results in the remaining three rows come from regressions that include all three treatments together.

Starting from the first two rows, OLS estimates in Columns (1) suggest high and positive partial correlation between average peer quality and students' KS3 achievements. The estimated coefficient is approximately 0.36 when only the average peer quality is entered in the regression, and it drops to 0.19 when the quality of top and bottom peers is further appended to the specification. This suggests that the tails of the ability distribution potentially capture most of the partial correlation between average peer ability and KS3 achievements.¹⁵ A similar picture emerges when looking at Columns (3) and (5), which display OLS estimates of the effect of top 5% and bottom 5% peers at schools: the estimated coefficient of good peers is large, between 0.75 and 0.33, while the estimated coefficient of bad peers is significantly negative and in the order of -0.6/-1.0.

However, a markedly different picture emerges when looking at Columns (2), (4) and (6). These comes from specifications that include pupil fixed-effects as described by Equation (1) in Section 3.1, and rely on within-pupil variation in age-14 test scores and peer quality to identify ability peer effects. Column (2) shows that the positive impact of average peer quality completely disappears upon inclusion of pupil fixed-effects: this is now estimated to be around 0.02, and not statistically different from zero. Similarly, Column (6) shows that the within-pupil estimates of the effect of the most academically talented peers are small and not statistically different from zero. Only the effect of the bottom 5% peers remains sizeable and significantly negative after including pupil fixed-effects. As shown in Column (4), this is estimated to be -0.12 in the first row, and -0.09 in the second row, where all three treatments are included simultaneously. Focusing on the latter, this is approximately one sixth of the corresponding OLS estimate. Although one reason why within-pupil estimates of peer effects might be smaller than OLS is because they net out overall effects that might arise through cross-subject interactions, this dramatic reduction is more likely due to the fact that within-pupil estimates control for pupil own unobserved average ability, unmeasured family background characteristics and school-by-cohort unobserved effects.

Nevertheless, pupil fixed-effects estimates presented in the first two rows are unconditional of KS2 achievements, and thus potentially contaminated by subject-specific pupil sorting. Therefore, in the last two rows of Table 3, we go on to include lagged test scores as an attempt to control for any residual pupil subject-specific ability and sorting. Note again that the specification in the third row presents estimates from specifications where we include pupils' own KS2 attainment in the same subject in interaction with subject dummies, while in the last row we include pupils' own KS2 test scores in the *same*-subject and *cross*-subject in interaction with subject effects. This 'control function approach' follows the strategy described in Section 3.2.

Comparing the second to the third and fourth rows, we find that OLS estimates of ability peer effects are now between 15% and 30% smaller than before. However, even when controlling for lagged test scores in the OLS specification in a very flexible way as in Row (4), we are unable to

¹⁵ To avoid double counting, we have also computed and experimented with measures of the average peer quality that exclude the top 5% and bottom 5% tails, and have come to identical conclusions.

reduce our estimate of the effect of peers' 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 and school-by-cohort unobservable shocks. On the other hand, the within-pupil estimates are essentially *unaffected* by the inclusion of pupils' age-11 test scores. The effects of the average peer quality and of the share of bright students remain small and insignificant. More interestingly, the effect of the bottom 5% peers only marginally drops to -0.091 (from -0.095), when we only include KS2 attainment in the same subject, and to -0.089, when we further include cross-subject lagged test scores.¹⁶ This finding is particularly reassuring especially considering that the same-subject lagged test score enters the within-pupil regressions with a large coefficient (of about 0.354, for example, in the third row), and is highly significant (t-statistics in excess of 230). In fact, the reason why lagged test scores hardly affect within-pupil estimates of effect of the bottom 5% new peers is that there is neither a sizeable nor a significant correlation between the within-student across-subject variation in age-11 achievements, and the variation in 'bad' peers' ability across subjects. Stated differently, conditional on pupil fixed-effects, the fraction of bottom 5% peers in one subject is completely balanced with respect to pupils' own age-11 test scores in that subject. To assess this more formally, we re-ran the regression in Equation (1) replacing age-14 with age-11 pupil test scores as the dependent variable. The coefficient on the fraction of bottom 5% peers was small at about -0.024 (with a standard error of 0.013), and not significant at conventional levels. On the other hand we found some degree of positive selection on the average peer quality, and some negative selection on the fraction of top 5%. This however does not substantially affect our results, which are steadfastly anchored at zero as soon as we include pupil fixed-effects. All in all, these results suggest that within-pupil specifications effectively take care of the endogenous sorting of pupils and their peers into secondary education, and that any residual subject-specific sorting is too small to confound out estimates.

5.2 *Robustness checks to potential threats to identification*

In this section, we present a set of robustness checks that further support the causal interpretation of our findings. Results from these exercises are presented in Table 4. Throughout the table, estimates come from within-pupil specifications that control for same- and cross-subject KS2 test scores interacted with subject specific dummies as described by Equation (2). 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, 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

¹⁶ We have also tried some specifications where we further include age-7 test scores. These are available for only three out of four cohorts, and students are not tested in science at age 7, so that we had to impute test scores in this subject using the average between mathematics and English. Even then, our findings were fully confirmed, with no effects coming from average peer quality and top students, and strong negative (same size) effects from the fraction of bottom 5% peers.

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; business and enterprise; and arts. 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 identical to those discussed in Table 3 for all peer quality measures.

Next, in the second row of the table, we look into whether results are driven by the fact that the school is above capacity (over-subscribed) or not (at capacity or under-subscribed). 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 can 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. To allay these concerns, Row (2) of Table 4 presents results obtained excluding over-capacity schools (accounting for approximately 40% of pupils in non-specialist schools). The within-pupil estimates of the effects of peers’ quality are similar to those obtained before, in particular for the impact of the fraction of bottom 5% new peers, which is now slightly larger at -0.100 (s.e. 0.040). Results (not tabulated, but available upon requests) further show that our findings are similar for non-specialist secular schools and non-specialist schools with a religious affiliation. All in all, the evidence suggests that neither *school-side* selection of pupils with unobservables potentially correlated with ability in a given subject, nor other school institutional features are driving our main results.

Another robustness check assesses whether *parental choice* of schools with an ‘expertise’ in a given subject might confound our estimates of peer effects. To do so, we 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 identify two groups of students: (i) those 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 (ii) those 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 as ‘sorted’ and ‘mixed’ pupils, respectively.¹⁷ We then re-run our analysis only including ‘mixed’ students to understand whether our results are driven by sorting of pupils with similar unobservables that are conducive to excellence in subject q (e.g. English) in the same school. Results from this exercise are reported in Row (3) of Table 4 and support our previous findings. Even

¹⁷ Note that peers’ excellence in a subject is defined using new peers’ average KS2 test scores. However, our results are unaffected if we use the fraction of new peers in the top 5% of the ability distribution.

when considering only ‘mixed’ pupils, we find no significant effects from peers of average quality and from the fraction of new peers in the top 5% of the ability distribution. On the other hand, we still find a sizeable and statistically significant negative effect from the bottom 5% peers. The estimated impact is at -0.100 (s.e. 0.034), which fully confirm our results so far.

To provide further evidence of the validity of our specifications, we next perform a robustness check based on replicating our results for increasingly selected subsets of students with increasingly small within-pupil standard deviation of KS2 test scores across the three subjects. While we move towards more ‘limited’ samples, we reduce the possibility that there is 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 forced to become progressively close to zero. In the empirical application, we perform this exercise by selecting students with the within-pupil standard deviation of KS2 test scores across the three subjects below increasingly smaller thresholds (e.g. $s.d.\leq 4$, $s.d.\leq 3.5$, $s.d.\leq 4$, $s.d.\leq 3$, etc.).¹⁸ Following the reasoning in Altonji et al. (2005), any residual sorting on *unobservable* subject-specific attributes most likely tracks and is upward bounded by the amount of selection on *observable* subject-specific characteristics, in particular lagged tests scores. Thus, by focusing on progressively ‘limited’ samples of students with little or no within-pupil variation in age-11 achievements and by studying how our estimates of the ability peer effects change, we are able to assess whether any residual subject-specific sorting might bias our estimates.

We present our findings graphically in Figures 1, where we focus on the bottom 5% new peers.¹⁹ The plots present regression coefficients and 95% confidence intervals (standard errors clustered at the school level) coming from 23 different regressions estimated separately for progressively small subsets of pupils with variation across subject in KS2 test scores falling below predefined thresholds of the within-pupil standard deviation of age-11 attainments. These spanned the interval $s.d.\leq 3$ to $s.d.\leq 11.5$, in steps of 0.5, and then $std.dev.\leq 15$; $std.dev.\leq 17.5$; $std.dev.\leq 23$; $std.dev.\leq 26$; and full sample. Note also that the estimates presented in the top panel come from specifications as in Equation (1), where the dependent variable is pupil age-11 achievement, and therefore present the balancing of this treatment with respect to pupils’ KS2 test scores. On the other hand, the estimates displayed in the bottom panel are obtained from specifications as in Equation (2), and thus present how our the treatment effect varies across different groups of pupils.

The top panel shows that the share of ‘bad’ peers is not significantly related to within-pupil variation in KS2 test scores almost throughout the various sub-samples. Even when this relation reaches some statistical significance, the degree of unbalancing is very small. Expectedly, as we move to more restricted sample of pupils, the balancing gets closer to perfect with estimated coefficients of

¹⁸ Note that identifying the ‘limited’ samples 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 (2), and only keep pupils with a ‘close-enough’ match to themselves across subjects.

¹⁹ Results for the other two treatments lead us to identical conclusions. They are not reported for space reasons, but are available from the authors.

-0.005, -0.002 and exactly zero for pupils with $s.d. \leq 4$, $s.d. \leq 3.5$ and $s.d. \leq 3$, respectively. However, the most remarkable findings from this exercise appear in the bottom panel: even as we shrink the within-pupil standard deviation of KS2 test scores towards zero, we still find negative and significant estimates of the effect of the bottom 5% peers. More importantly, these estimates are stable at approximately -0.09 throughout the plot. For example, they take values of -0.084 (s.e. 0.032) and -0.109 (s.e. 0.040) for the sets of pupils with $s.d. \leq 11.5$ and $s.d. \leq 3$ respectively. Furthermore, the confidence intervals throughout the figure are largely overlapping, clearly allowing us to reject the hypothesis that the estimates are different.

In conclusion, this last piece of evidence reinforces our main finding (evident in Tables 3 and 4) that any residual subject-specific sorting based on unobservable considerations must be sufficiently small not to confound our estimates of the effect of peers' quality conditional on pupil fixed-effects. In fact, any bias due to confounding subject-specific unobservables should have a very special pattern so as to lead to the same or slightly larger point estimates of the effects of 'bad' peers in samples of pupils with progressively shrinking degrees of variation in lagged test scores. In particular, selection on unobservables should be uncorrelated or negatively related to lagged test scores in order to explain these results. This is highly implausible since KS2 test scores are reliable proxies of pupils' subject-specific abilities, and it is very likely that pupils with similar subject-specific abilities or preferences will sort in the same schools.

5.3 *Extending the group of bottom and top peers beyond the 5% threshold*

One issue that we have so far left un-assessed is our choice of the 5% threshold to define the very good and markedly poor peers. Different cut-off points could have been chosen, potentially affecting our results. In Figure 2, we tackle this issue directly by looking at whether peers in other parts of the ability distribution significantly affect pupils' age-14 cognitive outcomes. The figure presents treatment effect estimates and associated 95% confidence intervals for different measures of the bottom and top new peers, and coming from specifications as in Equation (2). 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 to 90%; 80 to 85% and 75 to 75%. Note that the sample of 'treated' pupils now only includes students in the range from 25th to the 75th percentiles of KS2 test scores.

Figure 2 reveals a markedly asymmetric pattern. All five bottom peer groups have a negative effect on other pupils, but this effect is clearly significant only for the first group, and it declines sharply in scale as we move away from the very bottom group. On the other hand, the effect of the top peers at school is small and insignificant throughout. This suggests that our choice of top 5% and bottom 5% peers is not arbitrary and provide clear evidence that: (i) it is only the very bottom 5% of new peers that are strongly and negatively associated with pupils' own age 14 test scores, and not 'bad' peers in other parts of the quality distribution; and (ii) that there is no evidence that 'good' peers in other parts of the ability distribution affect students' cognitive outcomes.

To conclude this section, we provide an assessment of the magnitude of the negative effect of the bottom 5% peer treatment based on the estimates presented in Table 3. To do so, we begin by scaling it according to the minimum and maximum values of the bottom treatment variable observed in the data, at zero and 20% respectively (see Table 2). A pupil who moves from 20% to 0% of the bottom quality peer group would experience an improvement of KS3 test score of about 1.8/2 percentiles, which amounts to 0.08/0.09 of the standard deviation of KS3 test score, or 0.16/0.17 if we consider the standard deviation of the within-pupil KS3 distribution. Note that these are rather sizeable experimental changes, as they correspond to about 20 standard deviation changes in the within-pupil peer quality distribution. More modest changes of a 10 percentage point decline in the share of weak peers would imply an improvement of around 0.08 of the within-pupil standard deviation in the KS3 distribution. Relative to other studies that focus on school inputs and interventions, our estimates of the effect of academically weak peers capture a medium-to-small sized effect. For example, Lavy (2009) estimates the effect of instructional time in secondary schools using the PISA 2006 data and reports an average effect for OECD countries of 0.15 of the within-pupil standard deviation of test scores across subjects for an additional hour of classroom instruction. These estimates imply that the ability peer effects that we estimate here for a 10 percentile decrease in the percentage of bad peers quality is equivalent to the effect of half an hour of weekly instruction time. 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. Finally, Bandiera et al. (2009) 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 0.11 of the within-pupil standard deviation of outcomes.

5.4. *Estimates of the peer effects in small schools*

We next turn to analyze whether the within-pupil estimates of the peer effects are significantly different in small schools. As explained in Section 3.3, the possibility that schools implement subject-specific ability grouping (setting) means that we might underestimate the full extent of the scale of peer effects. By focusing on smaller schools and analyzing how our estimates change, we can partly allay these concerns. This is because schools with a smaller pupil intake will have fewer classes. Therefore students will be more mixed with peers of heterogeneous abilities in smaller schools than pupils in larger ones, where more classes can be created to group students according to their abilities. Notice that these arrangements stems from the fact that schools receive funding based on pupil number and have clear incentives to run classes at maximum capacity (approximately 30/35 students).

To perform this check, we focus on schools with pupil intake below the median of the year-7 cohort-size distribution. Stated differently, we consider (approximately) the 50% smallest schools with incoming cohort size of at maximum 180 pupils, and with on average of 136 students. Results are

reported in the last three rows of Table 4. Rows (4) and (5) present estimates that come from specifications as detailed in Equations (1) and (2), respectively. Once more, we find that controlling for pupil age-11 test scores in a very flexible way does not affect the within-pupil estimates. These are still clearly zero for the effect of the average peer quality and negative significant at around -0.10 for the fraction of new peers in the bottom 5% of the ability distribution. On the other hand, we find more positive effects for the fraction of top 5% new peers. This points into the direction of better peers interacting more with regular students in smaller schools and exercising more positive externalities. However, note that neither of the estimates in Rows (4) and (5) is significant at conventional levels. To further explore this issue, we also looked at the results for the smallest 25% schools and found that the effect of the very bright new peers is in the order of 0.060, but still not statistically significant, with an associated standard error of 0.044 (on the other hand, the effect of the bottom 5% new peers rises slightly to -0.12 with a standard error of 0.052).

One further advantage of focusing on small schools is that their peers' subject-specific quality is more likely to display significant year-on-year variation due to random subject-specific cohort shocks (recall that general cohort-specific unobserved effects are account for by our regression). We can exploit this fact to further augment our specifications with school-by-subject fixed-effects that account for subject-specific school unobservables – such as teachers' expertise in a given field – which might drive pupils' and their peers' sorting. We estimate this specification using only the first and last cohort in our data in order to maximize the variation over time that we can exploit to estimate ability peer effects. Indeed, this approach is very demanding since conditional on pupil fixed-effects our data shows very little within-school-subject variation over time, in particular in terms of students' age-14 outcomes. This is because the 'spread' of pupils' KS3 test scores around their average is not significantly widening or vanishing over time within schools. This fact is perhaps not surprising given that we are considering standardized test scores and that schools' composition does not dramatically changes over four years. Even then, our results (presented in Row 6 of Table 4) broadly support our previous conclusions. The effects of the average peer quality and the fraction of top 5% new peers are still estimated to be small and insignificant. On the other hand, the peer effect from the very weak students is estimated to be a significant -0.070 (s.e. 0.021), only between 20-30% smaller than our main estimates.

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 looked for some related evidence by running regressions separately for couples of subjects, i.e. by pooling observations for: English and Mathematics only; English and Science only; and Mathematics and Science only. Results are not tabulated for space reasons, but are available from the authors.

Our previous findings for the average quality of peers and the fraction of top 5% new peers were confirmed for all pairs of subjects. On the other hand, we found stronger peer effects from students in

the bottom 5% of the ability distribution coming from the comparison of English with Mathematics and English with Science, than when only pooling Mathematics and Science. For the former two couples of subjects, estimates of effect of ‘bad’ peers were -0.102 (s.e. 0.059) and -0.116 (s.e. 0.057) respectively, whereas the comparison of Science and Mathematics yielded a smaller estimate of -0.049 (s.e. 0.040). This is perhaps unsurprising given that, 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. Moreover, the institutional details discussed above suggest 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, both the within-pupil variation of the peers that the student *actually* interacts with, and the within-pupil variation in age-14 test scores might be too small to identify a significant peer effect. All in all, however, we believe the findings presented in this section broadly 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 six 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 (as well as subject dummies).

These findings are reported in Table 5. The estimates presented in Column (1) reveal that the quality of average peers does not affect regular pupils’ age-14 test scores at any point of the ability distribution. On the other hand, Column (2) shows the negative effect of the bottom 5% new peers is roughly constant across various ability groups of regular students. In fact, there is some variation in the point estimates, with larger negative effects for pupils in the 50th to 80th percentiles of the ability

distribution (at around -0.11) and insignificant negative effects for the most able pupils (of about -0.04). However, an F-test on the hypothesis that all coefficients are equal clearly 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). On the one hand, they confirm our main finding, namely that there is no significant peer effect from very academically bright pupils on other regular students. An F-test for the joint significance of the treatment at the various parts of the ability distribution clearly accepts the null of no effect. On the other hand, while the impact of the top 5% peers is positive (insignificant) in the bottom two-thirds of the ability distribution of regular students, it turns negative (insignificant) for the most able pupils with average age-11 test scores between the 65th and 90th percentile. Consistently, an F-test on the null that all coefficients are equal rejects the hypothesis at the 10% level of confidence with a p-value of 0.080.

Since this finding is rather unexpected, we have assessed its robustness along a number of directions. For example, we have tested that it survives when we restrict our attention to pupils with less potential for subject-specific sorting, as identified by students with a limited standard deviation of KS2 across subjects (i.e. pupils with $s.d. \leq 3$; see the discussion in Section 5.3). Similarly, we have tested that this pattern is not driven by the inclusion of specialist schools or over-subscribed schools. Finally, another possible and rather mechanical explanation for why pupils who are good on average marginally suffer from having many top 5% peers might be related to mean-reversion. In general, average test scores reveal some mean reversion. Pupils in the 5th-20th percentile at KS2 experience a 4 percentile point average improvement in their average KS3 test score, while students in the 80th-95th KS2 percentile have an average 5.6 percentiles 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 80th-95th 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, we would expect this to affect both the top and the bottom of the ability distribution. However, we do not observe any 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. To shed further light on this issue, we formally checked whether the pure effect of belonging to the top-group in the average KS2 ability distribution (80th-95th percentile) is related to the KS3 outcomes of students, but failed to find any evidence. In a nutshell, mean reversion does not appear to be a likely explanation for these patterns. Anticipating our findings, we find that this result is completely driven by the negative and significant response of boys to a large fraction of top 5% peers, which become particularly strong for the most able male students. We carefully investigate these issues in the next section.

6.2. *Gender heterogeneity in treatment effects*

In this section, we analyze the heterogeneity of peer effects by gender. This is particularly interesting given that a growing body of evidence shows that girls are more affected than boys by education

inputs and intervention.²⁰ Moreover, peer effects might work in significantly different ways for male and female students during secondary education, a time when both the identification with and the social interactions between the two genders intensify. We report our first set of results in Table 6. The top panel looks at boys (Columns (1) and (2)) and girls (Columns (3) and (4)) separately, but pooling pupils of all abilities. The bottom panel of the table instead further ranks students by their KS2 average ability. More details about the specifications are provided in the note to the table. Note that all regressions further include the average quality of peers. However, since this treatment did not reveal any significantly heterogeneous pattern, we have decided not to tabulate these coefficients (results available upon request).

Results in the Panel A of the table show that the effect of the bottom 5% peers is negative and significant in both gender groups, although is it slightly 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 0.066 for girls, but negative for boys at -0.052, and significant at better than the 10% level (p-value: 0.068). These patterns are not easily explained by differential subject-specific sorting for boys and girls into schools with peers of different quality. In fact, we find no significant relation between the within-pupil across subject variation in age-11 achievement and the variation in the fraction of top 5% new peers in different subjects for boys, and a small negative relation for girls (with coefficient of -0.064 and a standard error of 0.015), indicating some degree of negative sorting for female students. This clearly suggests that selection can hardly be driving our results: if this was the case, we should find more positive effects for boys than for girls (unless selection occurs on subject-specific unobservables that are *negatively* correlated with age-11 test scores)²¹. Note that we also checked whether our results are driven by the inclusion of single-sex schools. These enroll approximately 2% of the boys in our sample, and slightly more than 4% of the female students. Although results obtained after excluding these pupils were slightly weaker, they provided a similar picture: the effect of the bottom 5% peers is negative for both boys and girls, but the effect of the most academically talented peers is positive for female students and negative for males.

To shed further light on these patterns, we next study the sign and size of ability peer effects separately of boys and girls, and in interaction with students' own ability. Results are presented in Panel B of Table 6, and replicate the structure of Table 5. For both boys and girls, we find that the effect of many 'bad' peers at school is relatively stable throughout the ability distribution of regular students. The negative impact of bottom 5% peers is slightly stronger for pupils in the 50th-80th percentile range of the ability distribution. However, there is little evidence that these differences are

²⁰ 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 (see Angrist and Lavy, 2009 and Angrist et al., 2009).

²¹ Results unconditional on pupil's age-11 test scores confirmed these heterogeneous patterns for boys and girls.

statistically significant: an F-test on the hypothesis that all coefficients are equal accepts the null with p-values of 0.2597 and 0.6809 for boys and girls, respectively.

A more interesting pattern of results emerges when we focus on the effect of the top 5% new peers. Looking at girls first, we find that the impact of academically bright peers is positive throughout the ability distribution, although this effect is more pronounced and statistically significant for female students with KS2 achievements below the median of the ability distribution. On the other hand, the impact of top 5% peers becomes smaller and loses significance for the most talented girls, in particular for those with age-11 achievement above the 80th percentile, where the estimated coefficient is small and insignificant at 0.011 (s.e. 0.039). In sharp contrast, we find that the impact of having many 'good' peers at school is negative for males throughout the ability distribution, although this adverse effect is only statistically significant for the most able boys. The estimated impact for males with average KS2 test scores in the 65th to 80th percentile is -0.079 (s.e. 0.037), and further increases to -0.096 (s.e. 0.043) for those in the 80th to 95th percentile bracket. Note that we checked once again whether our results are driven by mean-reversion or ceiling effects. However, this does not seem the case. 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. Thus, a natural conclusion is that these effects are 'real', and the main question is what could explain them.

One possible explanation is based on 'crowding-out' effects: 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 groups (65th-80th and 80th-95th percentiles) from advanced activities, such as Science and Mathematics 'clubs', or special field trips because of limited space available in such activities. To clarify this, consider that there usually is only a limited number of places available in top-tier activities/clubs 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 extra-activities 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 65th to 95th percentile of the ability distribution.

One implication of this line of reasoning is that these negative effects should be mitigated in smaller schools. In fact, in these 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 activities. To check for this possibility, we re-run the analysis displayed in the bottom panel of Table 6 on the sample that only includes the 50% smallest schools. Our findings show that for schools in the bottom half of the cohort-size distribution

the positive impact of the top 5% peers for girls is positive and roughly constant throughout the ability distribution of regular students. Moreover, the effect of good peers is larger than before for girls in the top one-third of the ability distribution at approximately 0.075, although this estimate is not statistically significant. As for boys, we still find that a large share of top 5% peers at school has a negative impact on regular students, although the effects are now insignificant throughout the ability distribution and smaller in the top percentiles, at approximately -0.052. All in all, the evidence suggests that a crowding-out explanation of our findings might bear some relevance. However, this hypothesis cannot easily account for the still markedly different results that we document for males and female. In conclusion, we cannot exclude other more subtle explanations discussed in the educational and psychological literature, for example "big-fish-small-pond" mechanisms, which could be more pronounced for male students (see Marsh, 2005).

To conclude this section, we look at whether peer effects for boys and girls differ according to the gender of their peers. To do so, we re-compute the fraction of top 5% and bottom 5% new peers separately for male and female students, and re-run regressions similar to those in Panel A of Table 6, but including: the fraction of top 5% boys; the fraction of bottom 5% boys; the fraction of top 5% girls; and the fraction of bottom 5% girls. The average quality of peers is controlled for in these regressions, but not split along the gender dimension. This is because we found little evidence that peers of average quality matter for age-14 test scores of boys and girls. Note that the fractions of bottom 5% and top 5% new peers are now computed on very small number of students. Therefore, the statistical significance of our results is less indicative than the sign and magnitude of the coefficients.

These findings are presented in Table 7. Panel A tabulates results for boys, whereas Panel B deals with girls. Considering first the effect of the bottom 5% students, we find that boys are similarly affected by bad peers of both genders. Although the point estimates are slightly different across peers' gender, a test on the equality of the two coefficients accepts the null. Moreover, the estimated effect sizes are very close, at 0.405 and 0.452 for male and female peers respectively. These capture the percentage effect of one within-pupil standard change in either treatment on the within-pupil standard deviation in age-14 test scores. As for girls, evidence suggests that they are negatively affected by academically weak peers of both genders, although the adverse impact of bad female peers is more marked. Even though an F-test on the equality of the two coefficients accepts the null, the effect size of the bottom 5% female peers is almost twice as big as the one for bad male peers.

At the opposite end of the ability spectrum, we find that boys react more negatively to a large share of academically bright male peers, with an estimated coefficient is -0.073 (s.e. 0.039) corresponding to an effect size of negative 0.600. On the other hand the coefficient on the proportion of outstanding female peers coefficient is -0.034 (s.e. 0.044), with an effect size of negative 0.294. Remarkably, the opposite is true for girls, who respond more positively to bright peers of the same gender. The estimated effect of the top 5% female peers on other girls is 0.077 (s.e. 0.043) with an associated effect size of 0.797, whereas the effect of top 5% boys is as small as 0.037 (s.e. 0.042) with an effect size of 0.286.

6.3 *Additional findings: heterogeneity by pupils' eligibility for free school meals*

In this section, we briefly examine the heterogeneity of ability peer effects by pupils' eligibility for free school meals (FSM), a proxy for family income. To do so, we follow that same approach we used to look at gender differences in treatment effects. Namely, we first look at estimates obtained by pooling pupils of all ability groups, and then further break down peer quality estimates by pupils' own ability. Results are not shown for space reasons, but are available upon request.

Broadly speaking, results do not highlight any significant heterogeneity. Irrespective of pupils' eligibility for FSM, the bottom 5% new peers have a large and significantly negative impact on students' KS3 attainments. This is estimated to be -0.11 (s.e. 0.040) for FSM-eligible students, and -0.08 (s.e. 0.035) for pupils from richer background. On the other hand, we find that the average quality of peers and the fraction of good peers at school do not have any significant effect on students' performance irrespective of their FSM status. Similarly, we find very little evidence of heterogeneous effects when we further allow our estimates to vary along the dimension of pupils' ability. The negative effect of bad peers is sizeable and significant throughout for pupils of all aptitudes and irrespective of their FSM eligibility, except for students with KS2 average test scores in the 80th-95th percentile bracket, where the estimated impact remains negative but turns insignificant. On the other hand, the percentage of top 5% new peers has no significant impact on students' achievements irrespective of their ability and eligibility for free meals. Finally, we do not detect any interesting pattern for the effect of average peer quality. All in all, we find no evidence of heterogeneous peer effects along the dimension of family income.

7. **Concluding remarks and some policy implications**

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. Our analysis is highly relevant because of its 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 (6th grade) and Key Stage 3 (9th grade). Additionally, our sample is large enough to allow us to recover a variety of estimates about the heterogeneity of our treatment effects.

From a methodological perspective, we view our main 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 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, school-by-cohort fixed effects and other cohort-specific unobserved shocks, as well as pupil ability that is constant across subjects. Our findings strongly suggests that the within-pupil specifications take care of most of the sorting of pupils and their peers into secondary education, and provide reliable causal estimates of ability peer effects. However, to further support this claim, we have provided an extensive battery of robustness checks and falsification exercises that lend additional credibility to the causal interpretation of our results.

In terms of findings, our results clearly show that a large fraction of 'bad' peers at school as identified by students in the bottom 5% of the ability distribution is detrimental to other pupils' learning. On the other hand, we uncover little evidence that the average peer quality and the fraction of very 'good' peers as identified by students in the top 5% of the ability distribution affects the educational outcomes of other pupils across the board. However, these findings mask a significant degree of heterogeneity along the gender dimension. One striking result is that the very brilliant peers at school negatively impact the academic performance of boys, and in particular those who are among the highest groups 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 most able ability girls among regular students at school benefit the least from these interactions.

More in details, we have shown that a 10 percentile decrease in the proportion of 'bad' peers at school implies an improvement of approximately 0.07/0.09 of the within-pupil standard deviation of age-14 test scores for both boys and girls. On the other hand, a 10 percentage point increase in the percentage of 'good' peers would imply an improvement of 0.06 in the within-pupil standard deviation of KS3 achievements for girls and a nearly symmetrical negative effect for boys of 0.05. These differences become more remarkable if we consider boys and girls of different abilities. For the most talented males, a 10 percentage point increase in the proportion of top 5% peers implies up to 0.09 decrease in the within-pupil standard deviation of age-14 test scores. In sharp contrast, this same increase would boost achievements by more than 0.11 for the least able girls.

These heterogeneous patterns allow us to perform some concluding thought-experiments. To begin with, 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 would unambiguously improve male students' KS3 achievements by about 0.22 of a standard deviation (0.13+0.09) if we consider the within-pupil dispersion of KS3 achievements. Interestingly, this effect is not dissimilar for the most and least able boys, and is only slightly larger than 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. On the other hand, our thought-experiment would give more heterogeneous results for girls. On average,

the shift would improve female students' age-14 achievements by about 0.06 of a (within-pupil) standard deviation. This overall positive effect is the sum of the positive impact of not interacting with academically weak peers (at +0.18) and the adverse effect of reduced interactions with the best peers (-0.12). However, this overall effect would turn negative for girls in the bottom half of the ability distribution, with regular students in the ability bracket of 20th-35th percentile losing out as much as 0.10 of a (within-pupil) standard deviation. At the other extreme, the most talented girls could gain more than 0.20 of a (within-pupil) standard deviation of age-14 achievements from being educated in homogeneous environments with negligible fractions of 'bad' and 'good' peers.

Another policy-relevant experiment would be to simulate the effects of tracking by grouping all students – including the bottom 5% and top 5% – into two classes perfectly segregated along the lines of student's ability. The first group would include pupils who are above the median of the ability distribution, and the second those below the median. In this case, the lower ability group will experience a doubling of the proportion of bottom 5% pupils, on average from 4% to 8%, and a decline of the proportion of top 5% pupils from about 4% to zero. For the high ability class, the opposite will occur as the proportion of top 5% pupils doubles to about 8% and the proportion of bottom 5% falls to zero. These shifts would unambiguously worsen students' KS3 achievements in the low ability group, with a negative impact of about -0.03 in the within-pupil standard deviation of KS3 for boys, and -0.06 (-0.04-0.02) for girls. On the other hand, the changes experienced in the high ability group would improve boys' KS3 achievements by at most 0.01 (0.03-0.02) of a within-pupil standard deviation of KS3, while girls would benefit by up to 0.06 (0.04+0.02).

Do our results lend *overall* support to tracking of students by ability? Besides any equity consideration, we have shown that there is no simple answer to this question from an efficiency-of-learning point of view. Making schools more homogeneous by excluding both very good and very bad peers would result in an overall improvement in students' performance because, in our full sample, we find no positive effects stemming from a large fraction of top 5% new peers and significantly negative effects from bad peers. However, as we have just shown, our results are clearly heterogeneous in relation to one pupils' ability and gender, and vary according to the exact details of the tracking-experiment being carried out. One fairly stable finding is that female students above the median of the ability distribution could significantly benefit from either form of tracking, although other groups can be net losers or gainers (or unaffected) depending on the precise nature of ability grouping. In conclusion, despite not giving a one-size-fit-all policy recommendation, we believe our findings 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 Anna, (2008): "Peer Effects and Human Capital Accumulation: The Externalities of ADD," NBER Working Paper No. 14354.
- Altonji, Joseph G., Todd E. Elder and Christopher R. Taber (2005): "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools," *Journal of Political Economy*, 113, 151-184.
- Ammermueller, Andreas and Jörn-Steffen Pischke (2009): "Peer Effects in European Primary Schools: Evidence from PIRLS," *Journal of Labor Economics*, 27, 315-348.
- Anderson, Michael L. (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, Joshua D. and Kevin Lang, (2004): "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *American Economic Review*, 94(5), 1613-1634.
- Angrist, Joshua D., Daniel Lang and Philip Oreopolous (2009): "Incentives and Services for College Achievement: Evidence from a Randomized Trial", *American Economic Journal: Applied Economics*, 1(1), 136-163.
- Angrist, Joshua D. and Victor Lavy, (2009): "The Effect of High-Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99(4), 1384-1414.
- Arcidiacono, Peter and Sean Nicholson (2005): "Peer Effects in Medical School," *Journal of Public Economics*, 89(2-3), 327-350.
- Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster and Josh Kinsler (2009): "Estimating Spillovers with Panel Data, with and Application to the Classroom," mimeo, Duke University.
- Bandiera, Oriana , Iwan Barankay, Imran Rasul (forthcoming): "Social Incentives in the Workplace," *Review of Economic Studies*.
- Bandiera, Oriana, Valentino Larcinese and Imran Rasul (2009): "Heterogeneous Class Size Effects: New Evidence from a Panel of University Students," IZA Discussion Paper No. 4496.
- Becker, Gary S. (1974): "A Theory of Social Interactions," *Journal of Political Economy*, 82(6), 1063-1093.
- Bifulco, Robert, Jason M. Fletcher and Stephen L. Ross (2008): "The Effect of Classmate Characteristics on Individual Outcomes: Evidence from the Add Health," University of Connecticut, Department of Economics Working Paper No. 2008-21.
- Black, Sandra E. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114(2), 577-99.
- Black, Sandra E., Paul J. Devereux and Kjell G. Salvanes (2009): "Long After They're Gone? The Effects of Peers on Outcomes of Young Adults", paper presented at the CEPR/IZA European Summer Symposium in Labour Economics (ESSLE), Buch/Ameersee.
- Burke, Mary A. and Tim R. Sass (2008): "Classroom Peer Effects and Student Achievement," Federal Reserve Bank of Boston Working Paper No. 08-5.
- Carrell, Scott E. and Mark L. Hoekstra (2008): "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," NBER Working Paper No. 14246.
- Coleman, James S. (1966): "Equality of Educational Opportunity," Washington D.C., U.S., Department of Health, Education and Welfare, Office of Education, OE-38001.
- De Giorgi, Giacomo, Michele Pellizzari and Silvia 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 No. 14948.
- DfES (2006): "Grouping Pupils for Success," *Primary and Secondary National Strategies*, Department for Education and Skills, London.
- Duflo Esther, Pascaline Dupas and Michael Kremer (2008): "Peer Effects and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," CEPR Discussion Paper No. DP7043.

- Ellison, Glenn and Ashley Swanson (2009): "The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions," NBER Working Paper No. 15238.
- Figlio, David N. (2007): "Boys Named Sue: Disruptive Children and Their Peers," *Education Finance and Policy*, 2(4), 376-394.
- Fryer, Roland G. and Steven D. Levitt (forthcoming): "An Empirical Analysis of the Gender Gap in Mathematics," *American Economic Journal: Applied Economics*.
- Gibbons, Stephen, Stephen Machin and Olmo Silva (2009): "Valuing School Quality Using Boundary Discontinuity Regressions," London School of Economics, mimeo.
- Gibbons, Stephen and Shqiponja Telhaj (2008): "Peers and Achievement in England's Secondary Schools," SERC Discussion Paper 1.
- Glennerster, Howard (1991): "Quasi-Markets for Education," *Economic Journal*, 101, 1268-76.
- Gould, Eric D., Victor Lavy and Daniele M. Paserman (2009a): "Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi-Experimental Evidence," *Economic Journal*, 119, 1243-1269.
- Gould Eric D., Victor Lavy, and Daniele M. Paserman (2009b): "Sixty Years after the Magic Carpet Ride: The Long-Run Effect of the Early Childhood Environment on Social and Economic Outcomes," NBER Working Paper No. 14884.
- Goux, Dominique and Eric Maurin (2007): "Close Neighbours Matter: Neighbourhood Effects on Early Performance at School," *Economic Journal*, 117, 1-24.
- Hanushek, Eric A., John F. Kain, Jacob M. Markman and Steven G. Rivkin (2003): "Does Peer Ability Affect Student Achievement?," *Journal of Applied Econometrics*, 18(5), 527-544.
- Hoxby, Carlaine M. (2000): "Peer Effects in the Classroom: Learning from Gender and Race Variation," NBER Working Paper No. 7867.
- Kane, Thomas J., Stephanie K. Riegg and Douglas O. Staiger (2006): "School Quality, Neighborhoods and Housing Prices," *American Law and Economics Review*, 8(2), 183-212.
- Katz, Lawrence F., Jeffrey R. Kling and Jeffrey B. Liebman (2001): "Moving to Opportunity in Boston: Early Results from a Randomized Mobility Experiment," *Quarterly Journal of Economics*, 116(2), 607-654.
- Kutnick, Peter, Judy Sebba, Peter Blatchford, Maurice Galton and Jo Thorp (2005): "The Effects of Pupil Grouping: Literature Review," Department for Education and Skills Research Report 688.
- Kutnick, Peter, Steve Hodgekinson, Judy Sebba, Sara Humphreys, Maurice Galton, Susan Steward, Peter Blatchford and Ed 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.
- Lavy, Victor (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* (EEEEPE) Network Final Conference.
- Lavy, Victor and Analia Schlosser (2008): "Mechanisms and Impacts of Gender Peer Effects at School," NBER Working Paper No. 13292.
- Lavy, Victor, Daniele M. Paserman and Analia Schlosser (2008): "Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom," NBER Working Paper No. 14415.
- Lefgren, Lars (2004): "Educational Peer Effects and the Chicago Public Schools," *Journal of Urban Economics*, 56(2), 169-191.
- Manski, Charles F. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60(3), 531-542.
- Marsh, Herbert W. (2005): "Big Fish Little Pond Effect on Academic Self-concept: Cross-cultural and Cross-Disciplinary Generalizability", paper presented at the AARE Conference Paper.

- Moffit, Robert A. (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, Thomas and Jacob Vigdor (2007): "Peer Effects in North Carolina Public Schools," in *Schools and the Equal Opportunity Problem*, Ludger Woessman and Paul E. Peterson (eds.), Cambridge MA: MIT, 73-102.
- Sacerdote, Bruce (2001): "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, 116(2), 681-704.
- Sanbonmatsu, Lisa, Jeffrey R. Kling, Greg J. Duncan and Jeanne Brooks-Gunn (2006): "Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment," *Journal of Human Resources*, XLI(4): 649-691.
- Vigdor, Jacob and Thomas Nechyba (2004): "Peer Effects in Elementary School: Learning from 'Apparent' Random Assignment," Duke University, mimeo.
- Zimmerman, David J. (2003): "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Review of Economics and Statistics*, 85(1), 9-23.
- West, Anne and Audrey 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%
<i>Panel A: Pupils’ outcomes</i>			
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)
<i>Panel B: Pupils’ characteristics</i>			
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)
<i>Panel C: School characteristics (Year 7)</i>			
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)
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 regular pupils: approximately 1,200,000. The sample of regular students only includes pupils with KS2 achievement in each subject above the 5th percentile and below the 95th percentile of KS2 cohort-specific national distribution. Number of pupils with at least one subject in top 5% ($\geq 95^{\text{th}}$ percentile of KS2 cohort-specific national distribution): approximately 170,000. Number of pupils with at least one subject in bottom 5% ($\leq 5^{\text{th}}$ percentile of KS2 cohort-specific national distribution): approximately 130,000. 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. Community schools include only secular comprehensive state schools. Religiously affiliated schools include only schools in the state sector with some religious affiliation. Fractions may not sum to 1; this is due to rounding or partially missing information.

Table 2 – Descriptive statistics of treatments: average KS2 achievements and percentages of pupils in top 5% and bottom 5% of KS2 ability distribution – *new peers* only

Variable	Mean	Std. dev.	Min	Max
<i>Panel A: Average KS2 percentile treatment (new peers)</i>				
Average peer achievement at KS2 in English	49.79	8.71	1	98
Average peer achievement at KS2 in Math	49.94	8.06	1	100
Average peer achievement at KS2 in Science	49.68	8.35	1	100
<i>Panel B: Top 5% treatment (new peers)</i>				
Percentage, top 5% in English	4.22	3.03	0	19.56
Percentage, top 5% in Maths	3.77	2.60	0	19.87
Percentage, top 5% in Science	3.91	2.75	0	19.86
<i>Panel C: Bottom 5% treatment (new peers)</i>				
Percentage, bottom 5% in English	3.79	2.78	0	19.30
Percentage, bottom 5% in Maths	3.81	2.67	0	19.86
Percentage, bottom 5% in Science	3.78	2.90	0	19.78
Percentage 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. New peers refers to students in Year 7 in a given cohort that do not come from the same primary school.

Table 3 – Impact of peer quality on KS3 educational attainments: main results

Dependent variable is:	<i>Average peer KS2</i>		<i>Percentage of bottom 5% pupils</i>		<i>Percentage of top 5% pupils</i>	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
KS3 percentiles, unconditional on KS2; treatments entered separately	0.359 (0.012)**	0.022 (0.012)	-0.958 (0.029)**	-0.120 (0.033)**	0.750 (0.028)**	0.003 (0.003)
KS3 percentiles, unconditional on KS2; all treatments together	0.191 (0.013)**	0.018 (0.013)	-0.592 (0.032)**	-0.095 (0.033)**	0.332 (0.029)**	-0.021 (0.026)
KS3 percentiles, controlling for KS2 same-subject interacted with subject dummies; all treatments together	0.161 (0.011)**	0.012 (0.012)	-0.566 (0.030)**	-0.091 (0.033)**	0.244 (0.027)**	0.005 (0.026)
KS3 percentiles, controlling for KS2 same- and cross-subject interacted with subject dummies; all treatments together	0.146 (0.011)**	0.010 (0.012)	-0.511 (0.030)**	-0.089 (0.033)**	0.227 (0.027)**	0.008 (0.026)

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. **: at least 1% significant. Treatment effects in the first row estimated from two different sets of regressions: one including the average peer achievement at KS2 only (Columns (1) and (2)); and one including the percentage of top 5% pupils and the percentage of bottom 5% pupils in the cohort only (Columns (3) to (6)). All other regressions include all three treatments together. The table displays the coefficients on treatments based on new peers only. All regressions control for quality of old peers, and include subject and subject-by-gender dummies. Pupil characteristics controlled for in Columns (1), (3) and (5); absorbed in Columns (2), (4) and (6). Number of observations: approx. 3,600,000 (1,200,000 pupils), in 2193 schools.

Table 4 – Impact of peer quality on KS3 educational attainments: robustness to potential threats to identification and results for small schools only

Dependent variable is:	Within-pupil estimates		
	Average peer KS2	Percentage of bottom 5% pupils	Percentage of top 5% pupils
	(2)	(4)	(6)
KS3 percentiles, controlling for KS2: excluding specialist schools	0.010 (0.013)	-0.091 (0.034)**	0.013 (0.027)
KS3 percentiles, controlling for KS2: undersubscribed schools (excluding specialist)	0.013 (0.017)	-0.100 (0.040)**	0.011 (0.037)
KS3 percentiles, controlling for KS2: sample of pupils whose best subject is different from the best subject of new peers (<i>mixed</i>)	0.008 (0.013)	-0.100 (0.034)**	0.014 (0.027)
KS3 percentiles, unconditional on KS2: pupils in 50% smallest schools	0.005 (0.013)	-0.109 (0.045)**	0.028 (0.038)
KS3 percentiles, controlling for KS2: pupils in 50% smallest schools	-0.002 (0.013)	-0.104 (0.044)**	0.063 (0.038)
KS3 percentiles, controlling for KS2: pupils in 50% smallest schools; including pupils fixed effects and school × subject fixed effects	0.003 (0.006)	-0.070 (0.021)**	0.006 (0.017)

Note: All specifications as in Row (4) of Table 3, except in Row (4) where the specification does not control for lagged test scores. Specification in Rows (6) further includes school-by-subject fixed effects. Specialist schools account for about 8.5% of the pupil sample. Undersubscribed schools enrol approximately 60% of pupils in non-specialist schools. Sample of pupils with different best subject from new peers in school account for about 60% of the full sample. Sample of pupils in 50% smallest schools includes pupils in schools with less than 181 students in the year 7 cohort (approx. 6 classes of max 30 students). Regression with school × subject fixed effect (Row (6)) only considers the first cohort (year 7 in 2002) and last cohort (year 7 in 2005). Standard error clustered at the school level, except Rows (6) where they are robust. **: at least 1% significant.

Table 5 – Impact of peer quality on KS3 attainments: by pupil’s ability

Dependent variable is: KS3, controlling for KS2	Within-pupil estimates		
	<i>Average peer KS2</i>	<i>Percentage of bottom 5% pupils</i>	<i>Percentage of top 5% pupils</i>
	(1)	(2)	(3)
Effect for percentiles 5-20	0.011 (0.012)	-0.081 (0.029)**	0.016 (0.026)
Effect for percentiles 20-35	0.010 (0.015)	-0.068 (0.035)*	0.044 (0.030)
Effect for percentiles 35-50	0.012 (0.015)	-0.074 (0.041) [§]	0.023 (0.032)
Effect for percentiles 50-65	0.011 (0.016)	-0.118 (0.043)**	0.020 (0.033)
Effect for percentiles 65-80	0.005 (0.015)	-0.114 (0.043)**	-0.027 (0.031)
Effect for percentiles 80-95	0.016 (0.017)	-0.038 (0.050)	-0.049 (0.032)
<i>F-Test: all coeffs. jointly equal to zero (p-value)</i>	<i>0.9317</i>	<i>0.0321</i>	<i>0.1311</i>
<i>F-Test: all coefficients are equal (p-value)</i>	<i>0.9816</i>	<i>0.2564</i>	<i>0.0801</i>

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. Treatment effects estimated from one single regression including all three treatments together. The table displays the coefficient on treatments based on new peers. All regressions control for the quality of old peers. 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*. Ability blocks are defined using original KS2 percentiles computed out of the cohort-specific national distribution. 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). Specifications further include subject and subject-by-gender dummies. Number of observations: approximately 3,600,000 (1,200,000 pupils), in 2193 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; [§]: at least 10% significant.

Table 6 – Impact of peer quality on KS3 attainments, by pupil’s ability and gender

Dependent variable is: KS3, controlling for KS2	Within-pupil estimates			
	Boys only		Girls only	
	<i>Percentage of bottom 5% pupils</i>	<i>Percentage of top 5% pupils</i>	<i>Percentage of bottom 5% pupils</i>	<i>Percentage of top 5% pupils</i>
	(1)	(2)	(3)	(4)
<i>Panel A: Pupils of ability pooled (overall effect)</i>				
Overall effect	-0.076 (0.035)*	-0.052 (0.028) [§]	-0.098 (0.037)**	0.066 (0.029)*
<i>Panel B: Ability blocks defined on original KS2 percentiles</i>				
Effect for percentiles 5-20	-0.093 (0.032)**	-0.013 (0.029)	-0.080 (0.038)*	0.066 (0.035) [§]
Effect for percentiles 20-35	-0.057 (0.039)	-0.037 (0.033)	-0.072 (0.044) [§]	0.126 (0.037)**
Effect for percentiles 35-50	-0.068 (0.046)	-0.059 (0.036)	-0.066 (0.047)	0.088 (0.039)*
Effect for percentiles 50-65	-0.106 (0.048)*	-0.036 (0.038)	-0.113 (0.050)*	0.062 (0.038) [§]
Effect for percentiles 65-80	-0.089 (0.051) [§]	-0.079 (0.037)*	-0.139 (0.050)**	0.023 (0.036)
Effect for percentiles 80-95	0.036 (0.065)	-0.096 (0.043)*	-0.116 (0.060)*	0.011 (0.039)
<i>F-Test: all coeff. jointly equal to zero (p-value)</i>	0.0425	0.2642	0.1042	0.0334
<i>F-Test: all coefficients are equal (p-value)</i>	0.2597	0.4281	0.6809	0.0766

Note: Specifications in Panel A as in Row (4) of Table 3; specifications in Panel B as in Table 5. Separate regressions run for boys and girls. Number of observations for boys: approx. 1,800,000 (600,000 pupils) in 2101 schools. Number of observations for girls: approx. 1,800,000 (600,000 pupils) in 2134 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; [§]: at least 10% significant.

Table 7 – Impact of peer quality on KS3 attainments: treatments separately defined by pupils’ gender

Dependent variable is:	Within-pupil estimates			
	Percentage of bottom 5% pupils		Percentage of top 5% pupils	
	Counting male pupils only	Counting female pupils only	Counting male pupils only	Counting female pupils only
	(1)	(2)	(3)	(4)
<i>Panel A: Boys only</i>				
KS3 percentiles, controlling for KS2	-0.065 (0.049)	-0.090 (0.060)	-0.073 (0.039) [§]	-0.034 (0.044)
<i>Effect size</i>	<i>0.405</i>	<i>0.452</i>	<i>0.600</i>	<i>0.294</i>
<i>F-Test: coefficients are equal (p-value)</i>	0.7685		0.5364	
<i>F-Test: coeffs. jointly equal to zero (p-value)</i>	0.0929		0.1121	
<i>Panel B: Girls only</i>				
KS3 percentiles, controlling for KS2	-0.068 (0.053)	-0.124 (0.058)*	0.037 (0.042)	0.077 (0.043) [§]
<i>Effect size</i>	<i>0.414</i>	<i>0.755</i>	<i>0.286</i>	<i>0.797</i>
<i>F-Test: coefficients are equal (p-value)</i>	0.4980		0.5259	
<i>F-Test: coeffs. jointly equal to zero (p-value)</i>	0.0303		0.1168	

Note: The table reports regression coefficients and standard errors in round brackets from regressions of the dependent variable on treatments. Treatment effects estimated from one single regression including both treatments. The table displays the coefficient on treatments based on new peers and computed separately for male and female pupils. All regressions control for the quality of old peers computed separately for male and female pupils, and for the average quality of new and old peers. Controls further include KS2 percentiles in same- and cross-subject in interaction with subject dummies included, as well as subject dummies. 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. Number of observations: approximately 1,800,000 (600,000 pupils) in each panel. Number of schools: 2101 in Panel A; 2134 in Panel B. Standard error clustered at the school level. *: at least 5% significant; §: at least 10% significant.

Appendix Table

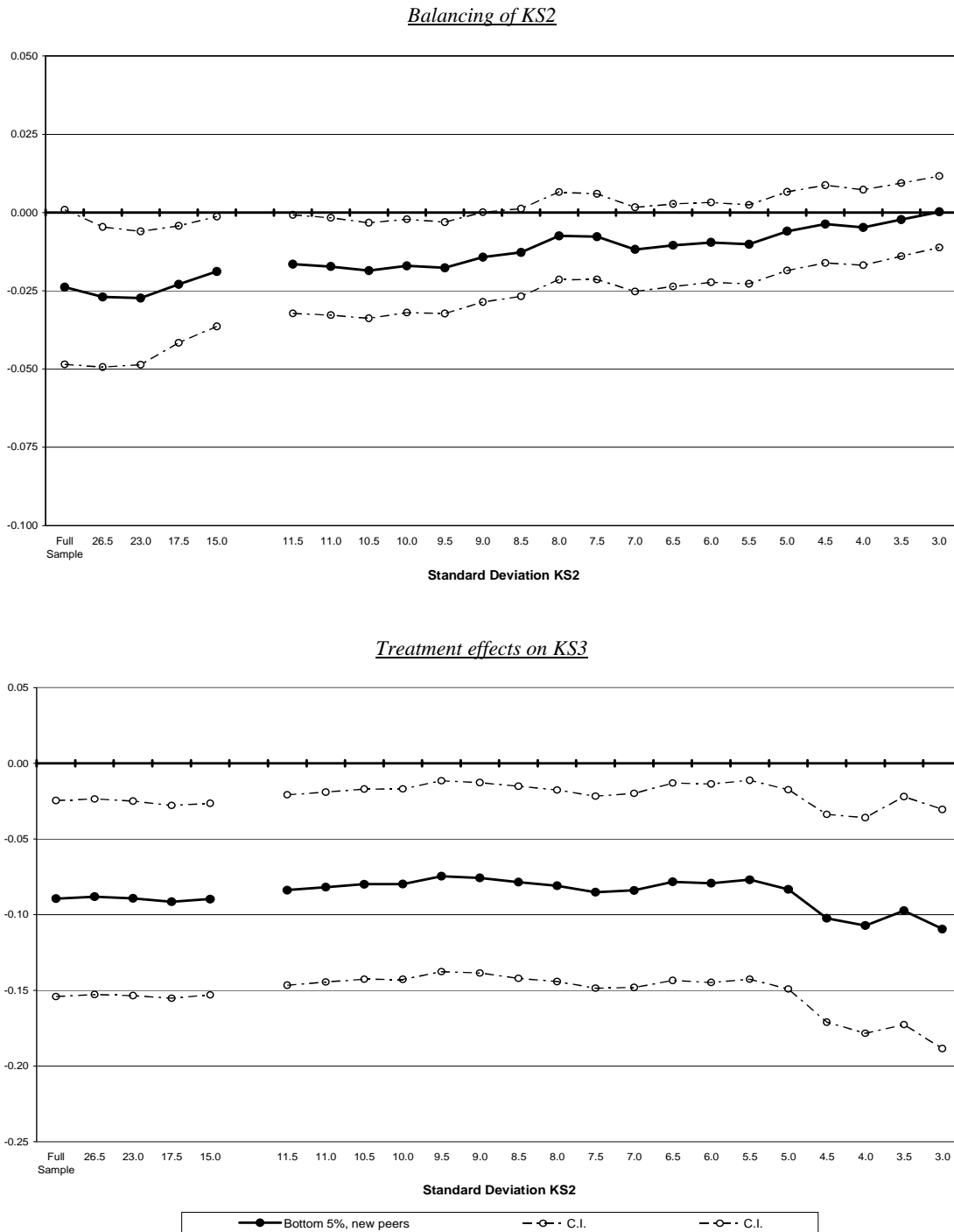
Appendix Table 1 – Within and between variation in pupil test scores and treatment measures

Variable:	Regular students				Sample including boys only				Sample including girls only			
	<i>Mean</i>	<i>Overall Std.dev.</i>	<i>Between Std.dev.</i>	<i>Within Std.dev.</i>	<i>Mean</i>	<i>Overall Std.dev.</i>	<i>Between Std.dev.</i>	<i>Within Std.dev.</i>	<i>Mean</i>	<i>Overall Std.dev.</i>	<i>Between Std.dev.</i>	<i>Within Std.dev.</i>
KS2 percentiles	49.19	24.31	21.15	11.98	48.71	24.38	21.10	12.21	49.66	24.22	21.19	11.73
KS3 percentiles	49.10	25.61	22.99	11.29	48.02	25.76	23.02	11.56	50.19	25.42	22.92	11.01
Average peer achievement at KS2	49.80	8.38	7.96	2.61	49.79	8.38	7.97	2.58	49.82	8.377	7.95	2.63
Percentage, bottom 5%	3.79	2.78	2.62	0.94	3.79	2.80	2.64	0.93	3.79	2.77	2.61	0.94
Percentage, top 5%	3.97	2.81	2.49	1.29	3.96	2.79	2.49	1.27	3.98	2.82	2.50	1.32

Note: Number of observations in the sample of regular students: approximately 3,600,000 corresponding to 1,200,000 pupils and 3 subjects. Number of observations in samples of boys and girls only: approximately 1,800,000 corresponding to 600,000 pupils and 3 subjects. Peer quality measures refer to new peers only.

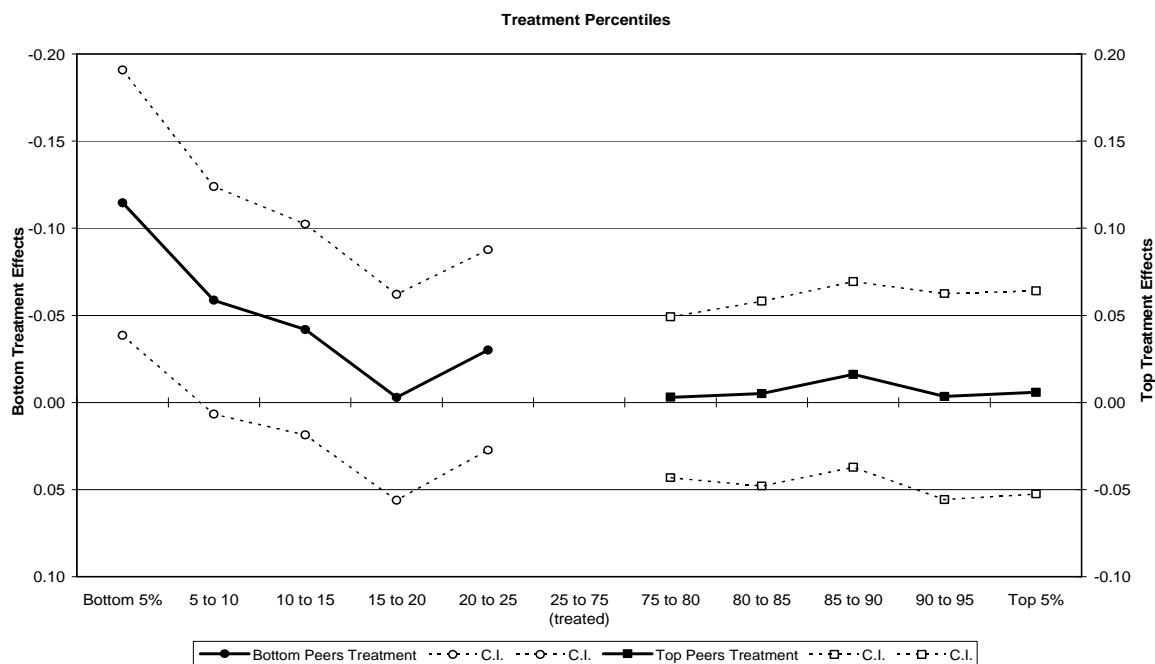
Figures

Figure 1 – Balancing and treatment effects of bottom 5% peers; by cumulative bands of the within-pupil standard deviation of KS2 attainments



Note: The figure plots regression coefficients and 95% confidence intervals (standard errors clustered at the school level) obtained regressing pupil KS2 achievements (top panels) and KS3 achievements (bottom panels) on the percentage of bottom 5% new peers. Regressions include: pupil fixed-effects; subject and subject-by-gender dummies; fraction of top 5% new peers and average quality of new peers; control for old peer quality. Regressions in the bottom panel further include pupil KS2 achievement in same- and cross -subject interacted with subject dummies. 23 different regressions were estimated over different cumulative bands of the standard deviation of KS2 attainments across subjects; these spanned the interval $\text{std.dev.} \leq 3$ to $\text{std.dev.} \leq 11.5$, in steps of 0.5, and then $\text{std.dev.} \leq 15$; $\text{std.dev.} \leq 17.5$; $\text{std.dev.} \leq 23$; $\text{std.dev.} \leq 26$; full sample.

Figure 2 – Treatment effects on KS3 percentiles; by different percentile cut-off points for top and bottom peers



Note: The figure plots regression coefficients and 95% 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. The regression further includes: pupil fixed-effects; pupil KS2 achievement in same- and cross-subject interacted with subject dummies; average new peer quality; controls for old peer quality; subject and subject-by-gender dummies. Treated pupils include students with KS2 achievements between 25th and 75th percentile of the cohort-specific distribution of KS2 for every subjects. Number of observations: approx. 2,580,000 (860,000 pupils) in 2193 schools.