The Good, The Bad and The Average: Evidence on Ability Peer Effects in Schools

Victor Lavy*, Olmo Silva** and Felix Weinhardt***

June 2011

Author for correspondence:
Victor Lavy
Department of Economics
The Hebrew University of Jerusalem
Mount Scopus, Jerusalem 91905, Israel
Tel.: 972-2-5883245
Email: msvictor@huji.ac.il

*Hebrew University of Jerusalem, University of Warwick, and NBER.
** London School of Economics, CEP and IZA.
***London School of Economics, CEP and IZA.

Acknowledgements: We thank the Center for Economic Performance (CEP) at the London School of Economics (LSE) for seed money for this project. We would also like to thank: Rebecca Allen, Josh Angrist, Kenneth Chay, Steve Gibbons, William R. Johnson, 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. Weinhardt also gratefully acknowledges ESRC PhD funding (Ref: ES/F022166/1). All remaining errors remain our own.

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

Abstract

We study ability peer effects in English secondary schools using data on four cohorts of pupils taking age-14 national tests and measuring peers' ability by prior achievements at age-11. Our identification is based on within-pupil regressions exploiting variation in achievements across three compulsory subjects tested both at age-14 and age-11. Using this novel strategy, we find significant and sizeable negative effects arising from bad peers at the bottom of the ability distribution, but little evidence that average peer quality and good peers matter. However, these results are heterogeneous, with girls benefiting from academically bright peers, and boys losing out.

1. Introduction

The estimation of peer effects in the classroom and at school has received intense attention in recent years. Several studies have presented convincing evidence about race, gender and immigrants' peer effects[1], but important questions about 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 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. 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 ability 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 (9^{th} grade) pupils entering secondary school in the academic years 2001/2002 to 2004/2005 and taking their age-14 national tests in 2003/2004-2006/2007. We link this information to data on pupils' prior achievement at age-11, when they took their end-of-primary education national tests, which we exploit to obtain pre-determined 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 majority of previous empirical evidence on ability peer effects in schools comes from studies that examine the effect of average background characteristics, such as parental schooling, race and ethnicity on students' outcomes (e.g. Hoxby, 2000 for the US, and Ammermueller and Pischke, 2009 for several European countries). A limitation of these studies is that they do not directly measure the academic ability of students' peers, but rely on socio-economic background characteristics as proxies for this. Additionally, our measures of peer quality are immune to refection problems (Manski, 1993) for two reasons. First, we identify peers' quality based on pupils' test scores at the end of primary education, before students have to change school and make a compulsory transition to the secondary phase. As a consequence of the large reshuffling of pupils in England during this transition, on average students meet more than 80% new peers at secondary school, i.e. students that do not come from the same primary. Secondly, 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 these respects, our strategy follows Gibbons and Telhaj (2008), also on English secondary schools. In our analysis, we

---

[1] See Angrist and Lang (2004) on peer effects through racial integration; Hoxby (2000) and Lavy and Schlosser (2011) on gender peer effects; and Gould et al. (2009) on the effect of immigrants on native students.

[2] Some exceptions are Sacerdote (2001) on ability peer effects among randomly paired roommates in university housing, and Carrell et al. (2009) on peer effects in squadrons at the US Air Force Academy.

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 some marked heterogeneity along the gender dimension. Indeed, we show that girls, especially those in the bottom half of the ability distribution, significantly benefit from interactions with very bright peers. In contrast, boys are marginally negatively affected by a larger proportion of academically outstanding peers at school. On the other hand, the negative effect of the very weak students does not significantly vary by the ability of regular students, nor along the gender dimension, and the effect of the average peer quality is estimated to be zero for boys and girls irrespective of their ability. Although we cannot pin down the exact mechanisms that give rise to these effects, we rationalize our findings by drawing on theoretical explanations and related evidence in the economics literature (e.g. Lazear, 2001; Hoxby and Weingarth, 2005; Jackson, 2009), as well as in the psychological and educational research (e.g. Cross and Madson, 1997; Eagly, 1978; Marsh, 2005).

Besides providing some novel insights about the nature of ability peer effects, our paper presents a new identification approach that allows us to improve on the (non-experimental) literature in the field and to identify the effects of peers' ability while minimizing 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 by ability. Using pupils' age-11 nationally standardized test scores as an indicator of ability we find that the average ability of peers and a pupil's own ability in secondary school are highly correlated. This is so despite the fact that most students have to change school when moving from primary to secondary education and that on average pupils meet more than 80% new peers. Similarly, there is a high correlation between pupils' and their peers' socioeconomic background characteristics, which is further evidence of sorting. More surprisingly, these correlations survive even when we look at the within-secondary-school variation over time of pupils' and their peers' ability and characteristic – i.e. conditional on secondary school fixed-effects.[4] This suggests that 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 strategies that rely on the randomness of peers' quality variation within-schools over time find little justification against this background.

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

[4] A similar result is documented by Gibbons and Telhaj (2008) and Black et al. (2010).

In order to overcome this selection problem, we rely on within-pupil regressions – i.e. on specifications including pupil fixed-effects – and 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. To the best of our knowledge, we are the first to use pupil fixed-effects and inter-subject differences in achievement to address identification issues of peer effects in schools.

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 influences. Additionally, we can partial out in a highly flexible 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. These include unobserved changes in school resources or head teachers, year-on-year variation in the student body's composition (e.g. the fraction of pupils from poor family background), as well as changes in the quality of primary schooling or childcare facilities. Given the evidence of year-on-year secondary school sorting highlighted above, controlling for these aspects seems particularly important.

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, there is neither a sizeable nor a significant correlation between the within-student across-subject variation in age-11 achievements – i.e. our measure for students' subject-specific academic ability – and the variation in peers' ability across subjects. Moreover, conditional on pupil fixed-effects, our results are virtually identical irrespective of whether or not we control for pupils' own age-11 test scores. Stated differently, specifications that include pupil fixed-effects effectively take care 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 to our core analysis. These include a set of regressions that focus on a sub-set of pupils with limited school choice from their place of residence, as well as results coming from specifications that further include school-by-subject effects to control for unobservable subject-specific school attributes. This additional evidence lends strong support to the causal interpretation of our results.

The rest of the paper is organized as follows. The next section reviews the literature on peer effects, while Section 3 presents our 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). 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.

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 a student's own quality and characteristics.[5] Failing to control for all observable and unobservable factors that determine individual sorting and achievements would result in biased estimates of 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., forthcoming; Carrell et al., 2009; De Giorgi et al., 2009; Gould et al., forthcoming), within-school random variation (Hoxby, 2000; Hanushek et al., 2003; Ammermueller and Pischke, 2009; Gould et al, 2009; Lavy and Schlosser, 2011), instrumental variables (Goux and Maurin, 2007) or sub-group re-assignments (Katz et al., 2001; Sanbonmatsu et al., 2006).[6] Only recently, Lavy et al. (forthcoming) and Duflo et al. (forthcoming) 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. (forthcoming) exploit 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.

[5] There is a well established literature on the link between school quality and house prices (e.g. Black, 1999; Gibbons et al., 2009; Kane et al., 2006), suggesting that pupils are segregated into different neighborhoods and schools by socio-economic status.

[6] Other examples include Aizer (2008), Bifulco et al. (2011), Burke and Sass (2008), Figlio (2007), Lefgren (2004), Nechyba and Vigdor (2007) and Vigdor and Nechyba (2004).

(forthcoming) present related evidence of significant negative effects of a high fraction of low ability students in the class 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 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. A similar approach has been previously used in Lavy (2010) to investigate the effect of instructional time on academic achievements; Bandiera et al. (2010) to study class size effects at university; and Dee (2007) to study the effect of teacher gender on students' attainments. As already mentioned, the within-student approach allows to control for pupil unobservable average ability, unmeasured family background influences, school-by-cohort fixed-effects and other more general cohort-specific shocks that are common to the three subjects. We believe this approach achieves a clean identification of the causal effect of peers' ability. The next section spells out in detail 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, potentially not fully observed. 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; Lavy et al., forthcoming; Lavy and Schlosser, 2011). 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 idiosyncratic and uncorrelated with students' potential outcomes and background.

In this paper, we suggest an alternative approach for overcoming the potential 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. Stated differently, we study whether pupils who have school peers that have 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:

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

Where *i* denotes pupils, *q* denotes subjects (English, Mathematics and Science), *s* denotes schools and *t* denotes pupils' cohort. A_{iqst} is an achievement measure for student *i* in subject *q* at school *s* in cohort *t*. In our analysis, we focus on test scores in the three compulsory subjects (English, Mathematics and Science) assessed at age-14 during the national tests; these are denoted in England as Key Stage 3 (KS3; more details in Section 4). Additionally, α_i is a student fixed-effect, β_q is a subject-specific effect, and γ_{st} is a school × cohort effect. We also include an interaction term between pupil 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, 2010, and Fryer and Levitt, forthcoming), and their potential effect on pupils' sorting into secondary schooling.[7] 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^h_{qst} and P^l_{qst} 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^h_{qst} and P^l_{qst} (this cut-off choice is not arbitrary; more details in the data and results sections). Finally, ε_{iqst} is an error term 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; and δ_2 and δ_3, which respectively measure the effect of the proportion of peers in the cohort who are in the top 5% and bottom 5% of the national distribution of KS2 test scores. As discussed above, we are interested in both the absolute and 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 and unmeasured family background characteristics.

[7] 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).

Moreover, this specification allows to partial out in a very flexible way school-by-cohort fixed-effects, such as unobserved changes in school resources or head teachers and year-on-year variation in the student body's composition (e.g. the proportion of students eligible for free school meals), as well as other cohort-specific unobserved shocks (e.g. changes in the quality of primary schooling or childcare facilities) that affect pupils' outcomes and peers' quality similarly across the three subjects. This seems particularly important given the issues discussed in Arcidiacono et al. (2009) and the evidence discussed in the Introduction of a significant correlation between pupils' characteristics and ability and the characteristics and ability of their peers even conditional on secondary school fixed-effects.

Before moving on, two remarks are worth being made. First, one necessary assumption for our identification strategy is that peer effects are the same for all three subjects; stated differently, we cannot interact the δ parameters with β_q in Equation (1). Although this restriction does not seem untenable, in the analysis that follows we will provide 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.

3.2. Dealing with potential threats to identification

Although the strategy described so far allows us to control for pupils' average ability across subjects, one concern is the possibility that 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 first approach to account for such residual sorting is to control for pupils' KS2 test scores in the within-pupil estimation. The underlying assumption is that lagged test scores effectively capture any subject-specific abilities and there is no sorting based on unobserved factors that are not correlated with KS2 scores, so that within-subject peer assignment is as good as random conditional on primary school test 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 age-14 English test score for pupil *i* controlling for 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 age-11 attainments in Mathematics and Science). This allows to partial out the effect of one pupil's own ability in a specific subject, as well as cross-subject effects. Additionally, we can interact lagged test scores with subject-specific dummies, so that age-11 achievements exhibit different effects on age-14 outcomes in different subjects. Under our most flexible and preferred specification, we estimate the following model:

$$A_{iqst} = \alpha_i + \beta_q + \gamma_{st} + \beta_q \times Gender + \delta_1 P_{qst} + \delta P^h_{qst} + \delta P^l_{qst} + \lambda_q a_{iqst-1} + \theta_q a_{iq(-1)st-1} + \kappa_q a_{iq(-2)st-1} + \varepsilon_{iqst} \qquad (2)$$

where a_{iqst-1} represents *same*-subject lagged test scores, $a_{iq(-1)st-1}$ and $a_{iq(-2)st-1}$ 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.[8]

Anticipating our analysis below, we find that results from within-pupil specifications are virtually identical whether or not we control for pupils' own age-11 test scores. This is explained by the fact that – as revealed by the placebo-type test carried out in Table 3 – 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 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. Nonetheless, we complement our core strategy with a set of robustness checks to assess the importance of subject-specific sorting. In particular, we restrict our analysis to a sub-set of pupils with limited school choice from their place of residence and for whom concerns about subject-specific considerations are mitigated. Even among these students, our findings hold completely unaffected.

A second source of concern is that, although the pupil fixed-effects strategy accounts for school-by-cohort unobserved shocks, it does not control for subject-specific school unobservables. This raises questions as to whether differences in pupils' attainments across subjects are driven by subject-specific differences in their peers' quality, or related to other factors such as school specialism in one area of the curriculum or teachers' subject-specific abilities.[9] In order to minimize these concerns, we exclude from our sample schools with a stated 'specialism' in a given subject. About 8.5% of secondary school students in England attend a specialist school, and some common areas of specialism include: language; mathematics and computing; science; technology; business and enterprise; and arts. Additionally, in some of our specifications we include school-by-subject fixed-effects – on top of pupil fixed-effects – to control for subject-specific school unobservables that are persistent over time. As detailed in the results section, these empirical models are very demanding in terms of the variation they exploit to identify the effects of peers' quality. Nevertheless, results from these specifications provide full 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 do not suffer from reflection problems. The longitudinal structure of the data that we use allows us to

---

[8] 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 outcomes.

[9] Carrell and West (2010) highlight the importance of teacher quality for university students' attainments.

link peers' age-11 test scores taken at the end of primary school (6th grade) to students' age-14 achievements four years later (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. In the national sample, on average 87% of pupil *i*'s peers at secondary school did not attend the same primary institution as student *i*, and therefore their age-11 test scores could not have been affected by this pupil. Following Gibbons and Telhaj (2008), in our analysis, we construct peer quality measures separately for new peers and old peers, and focus on the effect of new peers on pupil achievement. Nevertheless, we include measures of the quality of old peers in our empirical specifications to control for primary-school × cohort × subject effects that might persist on age-14 test scores and that are shared by pupils coming from the same primary school and cohort. Note that our estimates are not sensitive to the inclusion of these variables.

Two additional 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 age 14, 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. We do not see this as a particularly restrictive compromise since the majority of schools do not strictly group pupils with different subject-specific abilities into different classes at the early stages of secondary education (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. However, if a significant degree of subject-specific tracking takes place, grade-level peer quality measures might capture the peer quality actually experienced by pupils with some noise, thus leading to downward-biased estimates of the effect of peers' ability.[10] To minimize these issues, our main analysis focuses on the 50% smallest secondary schools in England, with a maximum grade-7 cohort-size of 180 students, and 135 7th grade-students on average. Small schools will have fewer classes since they receive funding based on pupil numbers and have clear incentives to run classes at maximum capacity (approximately 30/35 students). This implies that students will be more mixed with peers of heterogeneous abilities in small schools than in larger ones, where more classes can be created to group students according to their abilities. However, to further assess the importance of these issues, we will investigate whether our results change when we focus on a sub-set of even smaller secondary schools (average cohort size 120), as opposed to all schools in England irrespective of their size.

---

[10] Note however that even having access to information about class identifiers would not solve this issue if students can choose their networks – and thus their peers – within classes or outside of these. On the other hand, our study does not suffer from measurement error due to incomplete information on pupils' schoolmates as in Ammermueller and Pischke (2009).

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

Indeed, since the Education Reform Act of 1988, the 'choice model' of school provision has been progressively extended in the state-school system in England (Glennerster, 1991). In this setting, pupils can attend any under-subscribed school regardless of where they live and parental preference is the deciding factor. All Local Education Authorities (LEAs) and schools must organize their admissions arrangements in accordance with the current statutory Governmental Admissions Code of Practice. The guiding principle of this document is that parental choice should be the first consideration when ranking applications to schools. However, if the number of applicants exceeds the number of available places, other criteria which are not discriminatory, do not involve selection by ability and can be clearly assessed by parents, can be used to prioritize applicants. These vary in detail, but preference is usually given first to children with special educational needs, next to children with siblings in the school and to those children who live closest. For Faith schools, regular attendance at 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 applications, it has long been suspected that schools 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, at both KS2 and KS3 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 peers' quality as measured by their KS2.

Finally, regarding the organization of teaching and class formation, two important issues are worth mentioning. First, the notion of 'class' is a rather hollow one in English secondary schools since students are grouped with different pupils for different subjects. A second important aspect that – at least nominally – characterizes English secondary education is the practice of 'ability setting', i.e. subject-specific tracking. 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 topic.[11] However, despite some support from the central Government, the practice of ability setting is not liked nor supported by teachers, and as a result has not been fully adopted by secondary schools. Data collected by the Office for Standards in Education (OFSTED) in 2001-2002 from inspection of 566 secondary schools (cited in Kutnick et al., 2006) shows that only about 26% of these had subject-specific setting from 7th grade, with the percentage increasing to around 40% in 9th grade. Further evidence is provided by Kutnick et al. (2006), who gathered data for a small set of medium-to-large sized secondary schools with on average around 200 students in grade 7. This study shows that only around 50% of the schools had ability sets for Mathematics from 7th grade, with the figures being substantially lower for English and Science, respectively at 34% and 44%. Although these numbers increase as students reach 9th grade, subject-specific setting remains far from universal, with the figures being 46%, 59% and 80% in English, Science and Mathematics respectively. In conclusion, two features emerge from this 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. Finally, recall that our analysis focuses on the 50% smallest secondary schools in England – with on average 65 pupils (i.e. two classes) less than those sampled by Kutnick et al. (2006). This further minimizing issues due to subject-specific ability setting.

4.2. Data construction

The UK's Department for Children, Schools and Families (DCSF) collects a variety of data on all pupils and all schools in state education.[12] This is because the pupil assessment system is used to publish school performance tables and 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, from 2002 the DCSF has 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. PLASC is integrated with the pupil's assessment records in the National Pupil Database (NPD), giving a large and detailed dataset on pupil

[11] Subject-specific ability is often gauged using end-of-primary education KS2 test scores. However, these are only available to schools several months after they have admitted pupils and teachers have some discretion in determining the ability set that is most appropriate for their students in different subjects (see DfES, 2006).

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

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 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 to follow the academic careers of four cohorts of children from age-11 (6^{th} grade) through to age-14 (9^{th} grade), and to join this information to PLASC data for every year of secondary schooling (7^{th} to 9^{th} grade). The four cohorts that we use include pupils who finished primary education in the academic years 2000/2001 to 2003/2004, entered secondary school in 2001/2002 to 2004/2005, and sat for their KS3 exams in 2003/2004 to 2006/2007. We use information on these four cohorts 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. This is crucial to our analysis since we want to measure peer exposure at the beginning of secondary schooling (in 7^{th} grade), and not after three years (in 9^{th} grade). The data also allows us to gather information about the primary schools 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 capture peer effects coming from the 'very best' and the 'very worst' students at school, namely the fraction of peers in the grade below the 5^{th} percentile or above the 95^{th} 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 initial observations. We have also excluded selective schools (e.g. Grammar schools) from our analysis, as these can actively choose their pupils based on their ability (about 8% of our original sample). Furthermore, we drop schools where the fractions of pupils below the 5^{th} percentile or above the 95^{th} percentile of the cohort-specific KS2 national distribution do not exhibit any variation over the four years under analysis. These restrictions predominantly trim schools that have no students in either the top or bottom 5% of the ability distribution in any year in any subject, and would thus not contribute to the identification of peer effects. Since these constraints imply that we drop about 10% of the sample, we checked that our main results are not affected when we omit these restrictions. Finally, we focus on the 50% smallest secondary schools in England, with a maximum 7th grade cohort-size of 180 students, and 135 grade-7 students on average. Our final dataset includes a balanced panel of approximately 500,000 pupils for whom we can observe complete information in terms of KS2 and KS3 test scores, individual and family background characteristics, and both primary and secondary school level information from age-11 to age-14. In the next section, we present some descriptive statistics.

4.3. Some descriptive statistics

In Column (1) of Table 1 we present descriptive statistics for the main variables of interest for the sample of 'regular' students, defined as the set of pupils with age-11 test scores in the three subjects above the 5th percentile and below the 95th percentile of KS2 test score distribution. In the same table, we also presents descriptive statistics for pupils in either the top 5% or bottom 5% tails of the ability distribution – which we also label as 'treatments'. The regression analysis that follows is mostly based on the sample that includes all students, i.e. 'regular' pupils and the 'treatments'. However, we will also discuss some results based on the sample that excludes the top and bottom 5% pupils to keep the distinction between treated students and pupils that form our treatments clean.

In the top panel of the table we describe pupils' test scores at KS2 and KS3. Unsurprisingly, the first column shows that test score percentiles of regular students are centered just below 50, for all subjects and at both Key Stages. The correlations of pupils' KS2 test scores across subjects are 0.59 for English and Mathematics; 0.62 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 KS2 and KS3 test scores across the three subjects are respectively 11.8 and 10.9. 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. 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 analyzed the KS3 percentile ranking of pupils in the top 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%. In a nutshell, our 'good' and 'bad' peers are persistently among 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 the 'regular' sample is representative of the population of English secondary school pupils. On the other hand, pupils with at least one subject in the bottom 5% are less likely to have English as their first language 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.[13]

Finally, in Panel C we describe some school characteristics for the various sub-groups. Around 63% of all pupils attend Community schools, while about 25% of the pupils attend a religiously affiliated state-school. This figure is higher than in the national data (at around 16%), because religious schools tend to be of the smaller type that we sample here. Pupils with at least one subject in the top 5% of the ability distribution are less likely to attend a Community school, and more likely to be in a faith school, than pupils in the central part of the ability distribution and students with at least one subject in the bottom 5%. However, these differences are not remarkable.

In Table 2, we present some descriptive statistics of our 'treatments' for the new peers only. The top figures show that the median share of new schoolmates is 84%, although the distribution of new peers at school is right-skewed, with many more pupils facing almost 100% new schoolmates than at most 1%. The 25^{th} and 75^{th} percentiles of the distribution of new peers are 67% and 94%, respectively. Next, Panel A summarizes average peer quality (by-construction close to fifty for all subjects), whereas Panels B and C present descriptive statistics for our proxies for 'good' and 'bad' peers. Note that all peer quality measures display quite a wide range of variation, although this mainly captures 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. Finally, note that the incidence of pupils with at least one subject in the top 5% or the bottom

[13] Note that pupils in the bottom 5% of the KS2 distribution are more likely to change school between 7^{th} and 9^{th} grade. Additional results (not tabulated) also show that 'regular' students are less likely to change school during this period if they face more 'good' peers as well as more 'bad' peers. This advocates our use of peer quality measures based on schools attended in 7^{th} grade. On the other hand, we are not concerned with the overall effect of school mobility on pupil achievements as this is controlled for in our pupil fixed-effect strategy.

5% of the KS2 distribution is not concentrated in just a handful of schools: only four schools in our sample do not have at least some 'good' and some 'bad' peers in a given year. Moreover, the median, 10^{th} and 90^{th} percentiles of the school-by-year distribution of the percentage of very bright and very poor peers are respectively: 9.7%, 3.9% and 17.3%; and 7.3%, 2.6% and 15.1%.

5. Results

5.1. Effects of peers' ability: main findings

We begin the discussion of our results by presenting estimates of the impact of the peer quality on pupil outcomes at KS3 and controlling for any potential subject-specific sorting by including lagged test scores as discussed in Section 3.2. Results are reported in Panel A of Table 3. Columns (1) and (2) present OLS and within-pupil estimates of the effect of average peer quality. 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. These estimates 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. Finally, in the last row of Panel A, we include pupils' own KS2 test scores in the same-subject and cross-subject in interaction with subject effects. Note that the results in the first row are obtained from different regressions, where only one of the three peer quality measures is used as treatment. Results in the remaining rows instead come from regressions that include all treatments together. Finally, standard errors are clustered at the school level to allow for any degree of correlation in pupils' residuals across subjects, within schools and over cohorts.

Starting from the first two rows, OLS estimates in Column (1) show a high and positive partial correlation between average peer quality and students' KS3 achievements. The estimated coefficient is approximately 0.30 when only the average peer quality is entered in the regression, and drops to 0.12 when the quality of top and bottom peers is further appended to the specification. This suggests that the tails of the ability distribution capture most of the relation between average peer ability and KS3 achievements.[14] 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 on 'good' peers is large – between 0.83 and 0.46 – while the estimated association with 'bad' peers is significantly negative and in the order of -0.8/-1.1.

A markedly different picture emerges when looking at Columns (2), (4) and (6), where we report results from specifications that include pupil fixed-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 at most 0.01, and not statistically different from zero. Similarly, Column (6) shows that

[14] 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 similar conclusions.

the within-pupil estimates of the effect of the most academically talented peers are positive, but 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 -0.135 in the first row, and -0.124 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 background characteristics and school-by-cohort unobserved effects.

In the last two rows of Panel A of Table 3, we present estimates from specifications where we include lagged test scores as a way to control for any residual pupil subject-specific ability and sorting. Comparing the second row to the third and fourth, we find that the OLS estimates of ability peer effects are now between 10% and 50% 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 reduce our estimates of the effect of peers' quality to values close to the within-pupil estimates. This strongly speaks in favor of pupil fixed-effects regressions. On the other hand, the within-pupil estimates are essentially unaffected by the inclusion of pupils' age-11 test scores. The effect of the average peer quality remains small and insignificant, while the effect of the share of bright students increases from around 0.02 to approximately 0.04, but remains clearly insignificant. More interestingly, the effect of the bottom 5% peers only marginally drops to -0.120 from -0.124, when we include KS2 attainment.[15,16] 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 0.35 for example in the third row), and is highly significant.

In fact, the reason why the inclusion of lagged test scores hardly affects the within-pupil estimates of effect of peer quality is that there is neither a sizeable nor a significant correlation between the within-student across-subject variation in own age-11 achievements and the variation in peers' ability across subjects. Stated differently, conditional on pupil fixed-effects, peer quality in one subject is balanced with respect to pupils' own age-11 test scores in that subject. We show this formally in Panel B of Table 3, where we present results from regressions of one pupil's own age-11 test scores on the three peer quality measures (controlling for subject and subject-by-gender dummies). We label this regression analysis a 'placebo-type' test since we expect to find no relation if the

[15] We have also tried some specifications where we further include age-7 test scores. These are available for only three out of out four cohorts, Moreover, students are not tested in science at age 7 and 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.

[16] Note that the negative effect of 'bad' peers is slightly larger if we focus on students with a very high percentage of new peers at school. For example, considering the sample of pupils with at least 97% new peers (corresponding to the 10% percent of students with the largest fraction of new peers) we still find that only the fraction of bad peers has a significant impact, now estimated to be at -0.133 (s.e. 0.046).

variation in our 'treatments' is as good as random conditional on pupil fixed-effects. Columns (1), (3) and (5) present OLS estimates, whereas Columns (2), (4) and (6) come from the within-pupil specification detailed in Equation (1). OLS results show that unconditional on pupil fixed-effects there is a large and significant degree of sorting. For example, the association between pupils' own age-11 test scores and the fraction of bottom 5% peers is -0.40 and strongly significant. However, when we include pupil fixed-effects this relation drops by a factor of twenty to -0.021 (with a standard error of 0.017), and is not significant at conventional levels. Similarly, the within-pupil estimate in Column (2) shows that there is no significant relation between students' prior achievement in a given subject and the average peer quality in that subject. The estimated effect is as small as 0.012 and not statistically significant. Finally, the OLS estimate for the fraction of top 5% peers is 0.391 with a small standard error (0.017), suggesting large positive sorting. However, adding pupil fixed-effects eliminates this relation and reverses the sign of the placebo-test estimate to -0.034. Even though this coefficient is marginally significant, we regard it as spurious correlation. In fact, as we noted above, the estimated effect of the top 5% peers is not significantly changed when adding lagged test scores as controls. All in all, these findings 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.

Before moving on, note that the results so far come from regressions that include the top 5% and bottom 5% peers in the sample that we use to estimate the peer effects. However, as discussed above, an alternative would be to exclude the 'good' and 'bad' peers from the estimation sample in order to keep the distinction between treated pupils (i.e. the 'regular' students of Table 1) and 'treatments' clean. If we follow this approach, we find very similar results: our estimates of the effect of average peer quality and the top 5% peer are both small and insignificant at 0.000 (s.e. 0.014) and 0.054 (s.e. 0.040), respectively. On the other hand, the effect of the bottom 5% peers is a significant -0.128 (s.e. 0.047), slightly larger than our baseline estimate in Row (4), Panel A, Table 3.[17]

5.2 Robustness to potential threats to identification

In this section, we present a set of robustness checks that support the causal interpretation of our findings. Results are presented in Panel A of Table 4. 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' applications to schools. However, when schools are over-subscribed, they have some discretion in prioritizing pupils for admissions and once concern is that

[17] Regarding the effect of average peer quality being zero, we further looked into this issue by using the specification of Row (4), but including in the regression *only* the average peer quality variable. When doing this, the within-pupil estimates goes from 0.002 to 0.012 (but remains insignificant). This suggests that the reason why average peer quality does not have a sizeable impact when we include proxies for peers in the ability tails is that these capture most of the relevant 'empirical action', and not because we estimate net peer effects.

they might covertly select students with characteristics that are particularly suited to their teaching expertise and other infrastructures specific to one of the three core subjects under analysis. Note that 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 (1) in Panel A of Table 4 presents results obtained by excluding over-subscribed schools (accounting for approximately 35% of the pupils in our baseline sample). The estimates of the effects of peers' ability are similar to those obtained before, in particular for the impact of the fraction of bottom 5% peers, which is now slightly larger at -0.131 (s.e. 0.048). Further results (not tabulated, but available upon requests) also show that our findings are similar for secular schools and schools with a religious affiliation. All in all, this suggests that neither school-side selection of pupils with unobservables potentially correlated with ability in a given subject, nor other school institutional features drive our main results.

A second robustness check assesses whether parental choice of schools with an 'expertise' in a given subject might confound our estimates. To do so, we examine whether our findings are driven by sorting of students who 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, new peers excel in a different subject (either Mathematics or Science). We label these two groups as 'sorted' and 'mixed' pupils, respectively.[18] We then re-run our analysis including only 'mixed' students to understand whether our results are driven by sorting of pupils with unobservables that are conducive to excellence in subject q (e.g. English) in the same school. Results from this exercise are reported in Row (2) of Panel A of the table and support our previous findings. Even when considering only 'mixed' pupils, we find no significant effects from peers of average quality and 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 -0.120 (s.e. 0.046), which fully confirms our results so far.

To further assess the robustness of our findings against the possibility of subject-specific sorting, we perform two additional validity checks based on focusing on a subset of students with restricted 'school choice'. To carry out the first exercise, we exploit detailed geographical information on pupils' place of residence and location of the schools they attend, namely geo-coded postcodes with one-meter-precision geographical coordinates. Using this data, we start by calculating for each postcode of residence the median distance that pupils living in that postcode travel in order to attend their school. In our sample, this median home-school travel distance is on average 3km (or 1.9miles). For every pupil, we then count the number of schools other than the one currently attended that are within the median home-school travel distance for the postcode where he/she lives. We label this set of schools

[18] Note that peers' excellence in a subject is defined using new peers' average KS2 test scores. Our results are unaffected if we use the fraction of new peers in the top 5% of the ability distribution.

the pupil's 'choice set' (a similar approach was used in Gibbons et al., 2008). In our first robustness check, we focus on pupils with only one school within their 'choice set' and re-estimate Equation (2). This sub-group includes about 50 percent of the postcodes of residence in our sample, and 72 percent of the students. Results are reported in Row (3) of Table 4 and confirm our previous findings. The effect of bad peers is slightly smaller than previously found, but still sizeable at -0.099 and statistically significant. On the other hand, neither the average peer quality nor the fraction of peers in the top 5% of the ability distribution has a significant effect on students' age-14 attainments.

In our second robustness check, we take a coarser, but simpler approach and focus on students that have a limited number of schools within their Local Education Authority (LEA) of residence. As discussed in Burgess and Slater (2006), cross-LEAs school attendance is not predominant at secondary level (in particular outside London), and there is a strong presumption that pupils should attend a school within their LEA. Exploiting this intuition, we concentrate on students living in LEAs with the most restricted choice set and estimate Equation (2) on this subset of pupils. In particular, we focus on LEAs with less than 12 schools, which represent the median of the secondary school-per-LEA distribution. This sample includes approximately 140,000 students, or 28 percent of the full sample. Results from this exercise are reported in Row (4) of Table 4 and confirm the picture gathered so far. We still find that the average peer quality and the fraction of 'good' peers do not have significant effects. On the other hand, the impact of the fraction of peers in the bottom 5% of the ability distribution is larger than before at -0.151 and significant. Note that our results are not affected if we exclude all the London LEAs. All in all, these additional results reinforce the claim that our findings are causal and not driven by subject-specific sorting.

One final concern is that although the pupil fixed-effects account for school-by-cohort unobserved shocks they do not control for subject-specific school unobservables and therefore estimates of the effect of peers' quality might by confounded by other subject-specific school features, such as teachers' quality. As already stated, in order to minimize these concerns we have excluded from our sample schools with a stated 'specialism' in a given subject. To further allay these concerns, in Row (5) of Table 4, we include in our specifications school-by-subject fixed-effects – on top of pupil fixed-effects – to control for subject-specific school unobservables which are persistent over time. We estimate this specification using only the first and last cohorts in our data in order to maximize the variation over time that we can exploit to estimate peer effects. This approach is very 'demanding' since the identification lives off the variation in the 'spread' of pupils' KS3 test scores and peer quality measures across subject over-time within-schools, and this is not significantly widening or vanishing. This is perhaps not surprising given that we are considering standardized test scores and that schools' composition does not dramatically fluctuates over four years. Even then, our results broadly support our previous conclusions. The effect of the average peer quality is still estimated to be small positive and insignificant, whereas the effect of the top 5% peers turns small negative at -0.019, but clearly insignificant. On the other hand, the effect of 'bad' peers is still estimated to be a significant -0.086 (s.e. 0.019), only around 30% smaller than our main estimates.[19] We believe this reduction in size has more to do with attenuation biases, which are further exacerbated when adding an extra layer of fixed-effects (i.e. 3570 school-by-subject fixed-effects), than with other biases induced by subject-specific school unobservable attributes. Results presented in the next section back this intuition.

Before concluding, one possible concern is that our peer effect estimates might bundle together the effect of peers' ability with the effect of peers' background characteristics. Indeed, the descriptive statistics in Table 1 show that peers in the bottom 5% of the ability distribution are more likely to be eligible for free school meals and males, and less likely to be of White British origins; the opposite is true for pupils in the top 5% of the ability distribution. Nevertheless, these differences do not confound our estimates of ability peer effects. This is because the pupil fixed-effects regressions exploit within-pupil across-subject variation in peers' quality – not the overall variation across pupils – and the variation in the background characteristics of top 5% and bottom 5% peers across subjects is too small to have any significant effect on our findings.

5.3. Estimates of peer effects in alternative samples

Results presented so far focused on the 50% smallest schools in England. We argued that this is likely to mitigate attenuation biases arising from subject-specific tracking and the fact that we measure peer quality at the grade level, and not at the class level. In this section, we investigate how our results change when we consider some alternative samples. Results are presented in Panel B of Table 4, and come from within-pupil specifications as described by Equation (2).

To begin with, in Rows (1) and (2) of Panel B, we focus on the 33% smallest secondary schools in England. These schools have a grade-7 pupil intake of at most 158 students, and the average grade-7 cohort-size is 120 pupils (i.e. less than four classes of max 30/35 students). Results in the first row show that the effects of the average peer quality and the fraction of top 5% peers are still small and insignificant. On the other hand, the effect of the bottom 5% peers is around 15% larger than in our baseline specification at -0.141 (s.e. 0.052). Although this difference is neither large nor significant, this pattern suggests that our baseline estimates of the effect of 'bad' peers might be a lower bound to the effect of weak peers that we would be able to estimate in the complete absence of subject-specific tracking. On the other hand, there is no evidence that our main findings underestimate the effect of average peer quality and of 'good' peers.

In Row (2) of Panel B, we go one step further and include in our empirical model school-by-subject fixed-effects; this specification is comparable to the one in Row (5) of Panel A. Using this approach, we find that the fraction of 'bad' peers has a negative effect of -0.115 (s.e. 0.024), which is

[19] Note that when we include school-by-subject fixed-effects we do not cluster standard errors and simply use heteroskedasticity robust standard errors. This follows Angrist and Lavy (2009) who argue that school fixed-effects provide an alternative to clustering and absorb most of the within-school correlation in the error term. Nevertheless, even considering the standard errors obtained from our benchmark specification (see Row (4) of Table 3), our results on the effect of the bottom 5% peers would retain their significance.

approximately 18% smaller than the estimate we obtain without school-by-subject effects (Row (1) of Panel B). This attenuation is smaller than what we found using our baseline sample, and suggests that the reduction in the effect of the bottom 5% peers when including school-by-subject fixed effects is more likely due to measurement error – more pronounced in the sample of the 50% smallest schools – than to other subject-specific unobservables. On the other hand, the effect of the top 5% peers is now 0.044 and marginally significant. Note however that this estimate is above the one we obtain from the model that does not include school-by-subject fixed-effects. We take this as evidence that this finding is not robust and confirm our previous conclusion that the most academically talented peers do not have a significant impact on other students' achievements.

To conclude this section, we investigate what happens to our results if we avoid restricting the sample on the basis of school size and consider all secondary school (approximately 2300 schools, with an average grade-7 cohort-size of 183 pupils). Our findings are reported in Row (3) of Panel B, and confirm that the average peer quality and the fraction of top 5% peers do not significantly affect students' age-14 attainments. On the other hand, we still find that the fraction of bottom 5% peers has a significant and negative effect at -0.090 (s.e. 0.032). Although this estimate is lower than in our baseline sample, this finding shows that the negative impact of a large number of low achievers is not just a feature of small schools.[20] Furthermore, the pattern of variation documented in Panel B of the table provides evidence that subject-specific tracking in schools is not sufficiently widely-spread to significantly affect our estimates. This confirms the intuition gathered in Section 4.1. However, given the importance of these issues, we provide further evidence in the next section.

5.4. Additional findings: peer effects estimates by subject-couples

One of the underlying assumptions of our identification strategy is that peer effects are constant across different subjects. Although this assumption is difficult to test in our set-up, some of the studies that have previously investigated peer effects separately for Reading and Mathematics have found similar estimates (e.g. Vigdor and Nechyba, 2004; Hoxby and Weingarth, 2005), even though others have documented small differences across subjects (see the review in Epple and Romano, 2010). Similarly, there is mixed evidence about the differential impact of policy interventions and school resources on students' achievements in different subjects. See for example the discussion in Krueger (2003) in relation to class size.

From a conceptual point of view, one might argue that if different subjects rely on different skill sets and peer effects operate by changing pupils' skill sets, then we would expect to observe heterogeneous results for different subjects. On the other hand, peer effects from weak students of the type that we document here are more likely to emerge from pupils' behavior disrupting the normal functioning of class activities (as discussed in Lazear, 2001). In this case we would not necessarily expect heterogeneous effects for different subjects.

[20] Note that when using the school-by-subject fixed-effects (coupled with pupil fixed-effects) specification on the full sample of schools, we still find a significant impact of the bottom 5% peers, at -0.080 (s.e. 0.012).

To shed some light on this issue, we first ran OLS regressions separately by subjects and found that the estimated effects of average peer quality and 'bad' peers are virtually the same in the three subjects. These estimates are presented in the On-line Appendix Table 1. On the other hand, there is slightly more variation in the effect of top peers across the three subjects, but the confidence intervals of the various estimates are largely overlapping. Even though we know OLS estimates are biased, these findings are informative and suggest that our assumption is not unrealistic.

In another related check, we ran within-pupil 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 presented in the On-line Appendix Table 2. Our previous findings for the average peer quality and the fraction of top 5% 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 -0.142 (s.e. 0.078) and -0.179 (s.e. 0.074) respectively, whereas the comparison of Science and Mathematics yielded a much smaller estimate of -0.026 (s.e. 0.052). This pattern is perhaps unsurprising given that, as discussed in Section 4.1, the extent of subject-specific tracking is more prevalent in secondary schools for Mathematics (50% of the schools studied by Kutnick et al., 2006 have ability set from 7th grade), followed by Science (44%) and lastly by English (34%). As a result, when pooling the first two subjects only, students are not exposed to sufficient variation in peer quality to detect a significant effect.

Although one has to be cautious in interpreting the results presented in this section, we believe they provide some support for our assumption that peer effects are similar across subjects. Moreover, these findings further show that pupil sorting based on subject-specific considerations can hardly be driving our results. If this were the case, then estimates of the negative effects from 'bad' peers should remain strong and significant irrespective of whether we study subjects with little degree of tracking, as opposed to those with more significant ability setting.

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

One issue left un-assessed so far is our choice of the 5% threshold used to define 'good' and 'bad' peers. Different cut-off points could have been chosen, potentially affecting our results. In Figure 1, we tackle this issue by looking at whether peers in other parts of the ability distribution affect pupils' age-14 achievements. The figure presents estimates and 95% confidence intervals for different measures of the bottom and top new peers. For the bottom treatment, we define the following five groups: bottom 5%; 5 to 10%; 10 to 15%; 15 to 20% and 20 to 25%. For the top treatment, we define the following five groups: top 5%; 90 to 95%; 85 to 90%; 80 to 85% and 75 to 80%.

The figure 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 – twice turning negative – and insignificant throughout. This suggests that our choice of top 5% and bottom 5% peers is not arbitrary and provide evidence that: *(i)* it is mainly the very bottom 5% of new peers that is strongly and negatively associated with pupils' own age 14 test scores, and not 'bad' peers in other parts of the ability distribution; and *(ii)* 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 approximately 20% respectively (see Table 2). A pupil who moves from 20% to 0% of the bottom quality peer group would experience an improvement in KS3 test scores of about 2.4 percentiles, which amounts to around 9% of one standard deviation of KS3 test score distribution, or 22% if we consider the standard deviation of the within-pupil KS3 distribution. Note that these are sizeable changes, as they correspond to about 20 standard deviation changes in the within-pupil peer quality distribution. An alternative interesting benchmark looks at the improvement in peer quality associated with moving from an under- to an over-subscribed school. In this case, the fraction of 'bad' peers would decline by around 1.4 percentage points, which corresponds to an improvement in age-14 attainments of around 1.5% of one standard deviation in the within-pupil 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 (2010) estimates the effect of instructional time in secondary schools using the PISA 2006 data and reports an average effect for OECD countries of 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 is just below the effect of an extra hour of weekly instruction time. Another 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 student background measure of peer composition leads to a 17% standard deviation change in reading test scores of fourth graders. Finally, Bandiera et al. (2010) study class-size effects at university using a within-pupil specification similar to ours. Their results show that a one standard deviation change of the within-pupil class size distribution improves test scores by 11% of the within-pupil standard deviation of outcomes.

6. Allowing for Heterogeneous Effects

6.1. Heterogeneity by students' gender and ability

In this section, we test for the presence of heterogeneous effects along a number of dimensions. To begin with, 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.[21] 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 Panel A of Table 5. Columns (1) and (2) focus on boys while Columns (3) and (4) concentrate on girls. Results come from separate regressions for male and female students including all three peer quality measures. However, estimates for the average peer quality are not tabulated since they did not reveal any significant pattern. More details about the specifications are provided in the note to the table.

The results show that the effect of the bottom 5% peers is negative and significant in both gender groups, although it is slightly smaller for boys (at -0.108) than for girls (at -0.125). On the other hand, the effect of the top 5% peers is slightly negative for boys at -0.040 (s.e. 0.040), although not significant, but positive, significant and sizeable at 0.122 (s.e. 0.044) for girls. 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 a small negative insignificant relation between the within-pupil across-subject variation in age-11 achievement and the variation in the fraction of top 5% peers in different subjects for boys (with a coefficient of -0.025 and a standard error of 0.020), and a small negative borderline significant relation for girls (with coefficient of -0.040 and a standard error of 0.020). Moreover, results unconditional on prior achievement confirm this pattern: the effect of the top 5% peers for boys remains negative at -0.054 (s.e. 0.041), while the effect for girls is positive at 0.081 (s.e. 0.042). Differential sorting for boys and girls can thus hardly explain our results.[22]

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. For this purpose, we stratify the sample into six groups according to the distribution of pupils' average KS2 percentiles across the three subjects. The percentile-ranges that define the six non-overlapping groups are as follows: below 20 (includes the bottom 5% peers); 20-35; 35-50; 50-65; 65-80; and above 80 (includes the top 5% peers). The 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 which of the six KS2 ability groups a pupil belongs to. The effect of KS2 achievements in the *same*- and *cross*-subject is controlled for by interacting pupils' own KS2 percentiles with the dummies indicating his/her rank in the ability distribution as well as subject dummies. Our findings

[21] 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; Angrist et al., 2009).

[22] Note that we also checked whether our results are driven by the inclusion of single-sex schools. These enroll approximately 5% of the boys in our sample, and around 10% of the female students. Although results obtained after excluding these pupils were slightly weaker, they provided a similar picture.

are tabulated in Panel B of Table 5.[23] Once more, estimates for the average peer quality have not been tabulated since they were not significant. This is consistent with Hoxby and Wiengarth (2005), and speaks against 'linear-in-means' models which assume that pupil achievements are a linear function of the peers' average ability.

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 the bottom 5% peers is slightly stronger for pupils in the 50th-80th percentile range of the ability distribution, in particular for girls. However, there is no evidence that these differences are statistically significant: an F-test on the hypothesis that all coefficients are equal accepts the null with p-values of 0.5111 for boys and 0.4513 for girls. These results are consistent with a 'bad-apple'-type model of peer effects in which a small number of weak students (in a specific subject) adversely affects the learning of all other pupils throughout the ability distribution (Lazear, 2001; Hoxby and Weingarth, 2005**).**

A more heterogeneous pattern of results emerges when we focus on the effect of the top 5% peers. Looking at boys first, we find that the impact of the academically bright peers is negative although insignificant throughout the ability distribution. However, this effect is more pronounced and more precisely estimated (although still not significant at conventional levels) for male students with KS2 achievements above the median of the ability distribution, and in particular in the top 20 percentiles, where is stands at -0.067 (0.047). In contrast, we find that the impact of having many 'good' peers at school is positive for female students throughout the ability distribution, although this effect is more sizeable and statistically significant for girls in the bottom half of the ability distribution. The impact of top 5% peers becomes instead small 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 0.013 (s.e. 0.050). An F-test on the hypothesis that all coefficients are equal rejects the null with a p-value of 0.0045.

Since these patterns are rather unexpected, we have assessed their robustness along a number of directions. One mechanical explanation for why pupils who are good on average may marginally suffer (boys) or benefit less (girls) from having many top 5% peers might be related to mean-reversion. In general, average test scores reveal some mean reversion: pupils in the top 20% have on average a five-percentile deterioration in their average KS3. However, pupils within the same ability group, and in particular those above the 80th percentile, should all be similarly affected by mean-reversion irrespective of how many good peers they interact with. To shed light on this issue, we formally checked whether belonging to the top ability group is mechanically related to the KS3 outcomes, but failed to find any evidence. In a nutshell, mean reversion cannot explain these patterns.

Another 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

[23] On-line Appendix Table 3 presents results broken down by pupils' ability, but pooling boys and girls. Since most of the heterogeneity lies along the gender dimension, we do not discuss these findings here.

students who are in the next ability groups (65^{th}-80^{th} and 80^{th}-95^{th} 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 are usually 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 pupils 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 harming achievement. One implication of this line of reasoning is that these negative effects should be amplified in larger schools. In these schools the detrimental effect of having many top 5% peers might prevail, since there is potentially at the same time less room for interactions of pupils of different abilities and more 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 5 on the sample that includes all schools (not just the 50% smallest schools). In this case, we find that the negative impact of having many top 5% peers for the most able boys is larger than before and significant. In particular, this effect is -0.080 (s.e. 0.038) for males in the 65^{th}-80^{th} percentile of the KS2 ability distribution, and -0.075 (s.e. 0.031) for boys in the top 20 percentiles. Consistently, we also find that the positive impact of a larger fraction of top 5% peers for the most able girls is smaller when using the full set of schools. This effect is only 0.003 (s.e. 0.030) for the 20% most talented female students, and 0.027 (s.e. 0.037) for the next most able girls.

While this evidence suggests that a crowding-out explanation of our findings might bear some relevance, this hypothesis alone cannot easily account for the still markedly different results for boys and girls. Alternative more subtle explanations discussed in the educational and psychological literature should not be dismissed. In particular, research in these areas has highlighted marked gender differences in behavioral responses to settings that should lead to reciprocity, suggesting that females are more positively and cooperatively influenced by peers and social interactions than males (Cross and Madson, 1997; Eagly, 1978). Additionally, perverse 'big-fish-small-pond' effects have been shown to be more pronounced for males (see Marsh, 2005). Hoxby and Weingarth (2005) refer to these mechanisms as the 'Invidious Comparison Model', where the presence of high performing students depresses the performance of other students close-by in the ability distribution, possibly through envy, reduced motivation and self-esteem. On the other hand, a 'Shining Light'-type model of peer effects – where a few outstanding students can inspire all others to raise their achievement – might explain the response of female students to a larger fraction of 'good' peers (see again Hoxby and Weigarth, 2005).[24] In a very recent paper, Jackson (2009) finds gender heterogeneity in peer effects along the same lines discussed here.

6.2. Heterogeneity by peers' and students' gender

To conclude our investigation of gender heterogeneity, 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 5, 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, but not split along the gender dimension since we found little evidence that this matters 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 6. Panel A tabulates results for boys, whereas Panel B deals with girls.

Considering first results for boys (Panel A), we find that male students 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 0.605 and 0.687 for male and female peers, respectively. These capture the percentage effect of one within-pupil standard deviation change in either treatment on the within-pupil standard deviation in age-14 test scores. At the opposite end of the ability spectrum, we find that boys react negatively to a large share of academically bright male peers, with an estimated coefficient of -0.097 (s.e. 0.057) corresponding to an effect size of negative 0.827. On the other hand, boys are not negatively affected by a large fraction of outstanding female peers: the coefficient on the share of girls in the top 5% of the ability distribution is small and positive (but clearly insignificant) at 0.016 (s.e. 0.063), with an effect size of 0.143. This lends some support to explanations based on 'big-fish-small-pond' effects or 'invidious comparison' mechanisms (see Marsh, 2005; and Hoxby and Weingarth, 2005) where the presence of high performing male peers depresses the performance of other male students through envy, competition and reduced self-esteem.

As for girls, the evidence suggests that they are negatively affected by academically weak peers of both genders, although the adverse impact of bad female peers is significantly 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 – at 1.072 – is more than twice as large as the one for 'bad' male peers – at 0.470. On the other hand, the coefficient on the proportion of outstanding female peers is large and significant at 0.182 (s.e. 0.044) with an effect size of 1.978. Remarkably, this is not the case for outstanding male

[24] Note that we also tested whether the heterogeneity in peer's quality affects students' KS3 achievements. To do so, we included in our models an interaction term between the fraction of top 5% and the fraction of bottom 5% new peers. Although this variable enters our regressions with a negative sign for both boys and girls, the estimated effect is not significant. More importantly, including this term does not affect our main findings.

peers who exert no effect on the age-14 test score performance of female students. The estimated coefficient is small negative at -0.011 (s.e. 0.062), with an effect size 0.088. This is consistent with the economic and psychological literature (e.g. Anderson, 2008; Eagly, 1978) which shows that girls are more affected than boys by education inputs, and suggests that females are more positively and cooperatively influenced by peers and social interactions than males.

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

In this section, we briefly discuss the heterogeneity of ability peer effects by pupils' eligibility for free school meals (FSM), a proxy for family income. To analyze this issue, we followed the same approach used to look at gender differences in treatment effects. 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 fraction of 'bad' peers has a significantly negative impact on students' KS3 attainments. This is estimated to be -0.160 (s.e. 0.053) for FSM-eligible students and -0.102 (s.e. 0.046) for pupils from richer background. On the other hand, average peer quality and the fraction of 'good' peers do not have any significant effect on students' performance irrespective of their FSM status. Similarly, we find no evidence of heterogeneous effects when we allow our estimates to vary along the dimension of pupils' ability. In particular, 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 80^{th}-95^{th} percentile bracket, where the estimated impact remains negative, but turns insignificant. All in all, we find no evidence of heterogeneous peer effects along the dimension of family income.

7. Conclusions and Policy Implications

In this paper, we have estimated ability peer effects in schools using data for secondary schools in England for four cohorts of age-14 (9^{th} grade) pupils and measuring peers' quality by their academic ability as recorded by test scores at age-11 (6^{th} 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 cognitive outcomes at age 14. Our analysis is highly relevant because of its strong external validity: our results hold true in the nationally representative data which includes over 90 percent of four cohorts of pupils in England that transit from primary school through to the third year of secondary schooling, and sit for two crucial standardized national tests, namely the Key Stage 2 (6^{th} grade) and Key Stage 3 (9^{th} grade). Additionally, our sample is large enough to allow us to recover a variety of estimates about the heterogeneity of our treatment effects.

From a methodological perspective, we view our main contribution as offering a new approach to identification of peer effects based 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-effects estimation we are simultaneously able to control for family, school-by-cohort and other cohort-specific unobserved shocks, as well as for pupil ability that is constant across subjects. Our findings strongly suggests that the within-pupil specifications effectively take care 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 have provided an extensive battery of robustness checks that lend additional credibility to the causal interpretation of our results.

In terms of findings, our results 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 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 pupils across the board. However, these findings mask a significant degree of heterogeneity along the gender dimension, with girls significantly benefiting from the presence of very academically bright peers, and boys marginally losing out.

In more detail, our results imply that a 10 percentile decrease in the proportion of 'bad' peers at school is associated with an improvement of approximately 10-11% of the standard deviation of the within-pupil KS3 distribution 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 approximately 10% of the within-pupil standard deviation of KS3 achievements, but only for girls. This effect is more pronounced for female students at the bottom end of the ability distribution, where a 10 percentage point increase in the fraction of academically talented peers corresponds to an improvement of up to 23% of the within-pupil standard deviation of KS3 achievements. In marked contrast, male students stand to lose approximately 5% of the within-pupil standard deviation of KS3 from a 10 percentage point increase in the proportion of 'good' peers (even though this result is not significant), and around 9% if we consider the fraction of top 5% peers who are also males.

These heterogeneous patterns allow us to perform two concluding thought 'policy 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, i.e. a sort of tracking. This shift would unambiguously improve students' achievements by up to 20% of the standard deviation in the within-pupil distribution of KS3. Interestingly, this finding is not dissimilar from Carrell et al. (2011) who provide experimental evidence that 'middle ability' students assigned to more homogeneous classes experience improvements in their university GPA by around 11% of a standard deviation. However, our results mask some significant heterogeneity along the gender dimension. Whereas male students' achievements unambiguously improve by up to 25% of one standard deviation, this policy experiment would not improve female students' age-14 achievements on average: the positive impact of not interacting with academically weak peers – at 24% of a standard deviation – would be almost perfectly matched by the adverse effect of reduced interactions with the ‘good’ peers at -23%. Nevertheless, the effect would clearly turn negative for girls in the bottom half of the ability distribution, and substantially positive for the most talented girls.

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 around 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 class, with a negative impact of about -5% of a standard deviation in the within-pupil distribution of KS3 for boys and -12% for girls. On the other hand, the changes experienced in the high ability group would improve boys’ KS3 achievements by at most 5%, while girls would benefit by at least 5-6%. This policy experiment is not dissimilar from the randomized trial carried out by Duflo et al. (forthcoming), who document a 14% of a standard deviation improvement in the test scores of pupils in primary schools in Kenya after 18 months of random assignment to homogenous ‘tracked’ classes. Their findings hold true for boys and for girls, as well as for students assigned to the top and the bottom track. However, as the authors discuss in their work, these improvements are likely to arise because of changes in teachers’ behavior, in particular of those assigned to the bottom tier who closely tailored instruction time to class composition. On the other hand, our effects mainly arise from changes in peers’ quality across subjects holding teachers’ effort and instruction methods constant, which can perhaps explain the differences in our findings.

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: 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. Nevertheless, although we are fully aware of the difficulties of using reduced-form estimates to make out-of-sample policy predictions (see Carrell at al., 2011), we believe our data is rich enough – and our findings robust 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. Working Paper No. 14354, National Bureau of Economic Research, Cambridge, MA.

Ammermueller, Andreas and Jörn-Steffen Pischke. 2009. Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study. *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, Gigi Foster, Natalie Goodpaster and Josh Kinsler. 2009. Estimating Spillovers with Panel Data, with and Application to the Classroom. Unpublished Manuscript, Duke University, Durham, NC.

Bandiera, Oriana, Valentino Larcinese and Imran Rasul. 2010. Heterogeneous Class Size Effects: New Evidence from a Panel of University Students. *The Economic Journal*, 120:1365-1398.

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. 2011. The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, 3(1), 25-53.

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. 2010. Under pressure? The effect of peers on outcomes of young adults. Working Paper No. 16004, National Bureau of Economic Research, Cambridge, MA.

Burgess, Simon and Helen Slater. 2006. Using boundary changes to estimate the impact of school competition on test scores. Working Paper No. 158, Centre for Market and Public Organization, Bristol, UK.

Burke, Mary A. and Tim R. Sass. 2008. Classroom Peer Effects and Student Achievement. Working Paper No. 08-5, Federal Reserve Bank of Boston, MA.

Carrell, Scott E., Richard L. Fullerton and James, E. West. 2009. Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3), 439-464.

Carrell, Scott E., Bruce I. Sacerdote and James E. West. 2011. From Natural Variation to Optimal Policy? The Lucas Critique Meets Peer Effects. Working Paper No. 16865, National Bureau of Economic Research, Cambridge, MA.

Carrell, Scott E. and James E. West. 2010. Does Professor Quality Matter? Evidence from Random Assignment of Students to Professor. *Journal of Political Economy*, 118(3), 409-432.

Coleman, James S. 1966. Equality of Educational Opportunity. Department of Health, Education and Welfare, Office of Education, OE-38001, Washington D.C.

Cross, Susan and Laura Madson. 1997. Models of the Self: Self-Construals and Gender. *Psychological Bulletin*, 12, 5-37.

Dee, Thomas S. 2007. Teachers and Gender Gaps in Student Achievement. *Journal of Human Resources*, 42(3), 528-554.

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. Working Paper No. 14948, National Bureau of Economic Research, Cambridge, MA.

DfES. 2006. Grouping Pupils for Success. *Primary and Secondary National Strategies*, Department for Education and Skills, London, UK.

Duflo Esther, Pascaline Dupas and Michael Kremer. Forthcoming. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*.

Eagly, Alice H. 1978. Sex Differences in Influenceability. *Pshycological Bulletin*, 85, 86-116.

Ellison, Glenn and Ashley Swanson. 2010. The Gender Gap in Secondary School Mathematics at High Achievement Levels: Evidence from the American Mathematics Competitions. *Journal of Economic Perspectives*, 24(2), 109-128.

Epple, Dennis and Richard Romano. 2010. Peer Effects in Education: A Survey of the Theory and Evidence. Unpublished manuscript, Carnegie-Mellon University, Pittsburgh, PA.

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. 2008. Choice, Competition and Pupil Achievement. *Journal of the European Economic Association*, vol. 6(4), pp. 912-947.

Gibbons, Stephen, Stephen Machin and Olmo Silva. 2009. Valuing School Quality Using Boundary Discontinuity Regressions. Discussion Paper 18, Spatial Economics Research Centre, London, UK.

Gibbons, Stephen and Shqiponja Telhaj. 2008. Peers and Achievement in England's Secondary Schools. Discussion Paper 1, Spatial Economics Research Centre, London, UK.

Glennerster, Howard. 1991. Quasi-Markets for Education. *Economic Journal*, 101, 1268-76.

Gould, Eric D., Victor Lavy and Daniele M. Paserman. 2009. 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. Forthcoming. Sixty Years after the Magic Carpet Ride: The Long-Run Effect of the Early Childhood Environment on Social and Economic Outcomes. *Review of Economic Studies*.

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, Carloine M. 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation. Working Paper No. 7867, National Bureau of Economic Research, Cambridge, MA.

Hoxby, Caroline M., and Gretchen Weingarth. 2005. Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. Unpublished Manuscript, Harvard University, Cambridge, MA.

Jackson, Kirabo, C. 2009. Peer Quality or Input Quality? Evidence from Trinidad and Tobago. Unpublished Manuscript, Cornell University, Ithaca, NY.

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.

Krueger, Alan B. 2003. Economic Considerations and Class Size. *Economic Journal*, 113(485), F34-F63.

Kutnick, Peter, Steve Hodgekinson, Judy Sebba, Sara Humphreys, Maurice Galton, Susan Steward, Peter Bltachford and Ed Baines. 2006. Pupil Grouping Strategies and Practices at Key Stage 2 and 3: Case Studies of 24 Schools in England. Research Report 796, Department for Education and Skills, London, UK.

Lavy, Victor. 2010. Do Differences in School's Instruction Time Explain International Achievement Gaps in Math, Science, and Reading? Evidence from Developed and Developing Countries. Working Paper No. 16227, National Bureau of Economic Research, Cambridge, MA.

Lavy, Victor and Analia Schlosser. 2011. Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2), pp. 1-33.

Lavy, Victor, Daniele M. Paserman and Analia Schlosser. Forthcoming. Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom. *Economic Journal*.

Lazear, Edward P. 2001. Educational Production. *Quarterly Journal of Economics*, 116(3), 777-803.

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 Annual Conference, James Cook University, Smithfield, Australia.

Moffit, Robert A. 2001. Policy Interventions, Low-Level Equilibria, and Social Interactions. In *Social Dynamics*, ed. Steven N. Durlauf and Peyton H. Young, MIT Press, Cambridge, MA.

Nechyba, Thomas and Jacob Vigdor. 2007. Peer Effects in North Carolina Public Schools. In *Schools and the Equal Opportunity Problem*, ed. Ludger Woessman and Paul E. Peterson, MIT Press, Cambridge, MA.

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*, 41(4): 649-691.

Vigdor, Jacob and Thomas Nechyba. 2004. Peer Effects in Elementary School: Learning from 'Apparent' Random Assignment. Unpublished Manuscript, Duke University, Durham, NC.

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, London, UK.

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

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

Variable	'Regular' students	At least 1 subject top 5%	At least 1 subject bottom 5%
Panel A: Pupils' outcomes			
KS2 percentile, English	48.9 (24.3)	87.0 (14.8)	8.5 (12.1)
KS2 percentile, Mathematics	48.9 (24.3)	86.8 (14.1)	9.4 (13.3)
KS2 percentile, Science	48.6 (24.3)	87.7 (13.0)	10.8 (15.0)
KS3 percentile, English	49.0 (26.0)	81.3 (18.4)	15.4 (17.9)
KS3 percentile, Mathematics	49.1 (25.3)	84.4 (16.3)	14.9 (17.1)
KS3 percentile, Science	49.2 (25.5)	84.3 (16.2)	16.1 (17.7)
Panel B: Pupils' characteristics			
First language is English	0.93 (0.26)	0.95 (0.22)	0.89 (0.31)
Eligible for free school meals	0.14 (0.35)	0.05 (0.23)	0.30 (0.46)
Male	0.50 (0.50)	0.48 (0.50)	0.55 (0.50)
Changed school between 7th grade and KS3	0.20 (0.40)	0.20 (0.40)	0.22 (0.42)
Ethnicity: White British	0.85 (0.36)	0.88 (0.32)	0.81 (0.39)
Ethnicity: White other	0.02 (0.13)	0.02 (0.14)	0.02 (0.15)
Ethnicity: Asian	0.05 (0.21)	0.03 (0.18)	0.07 (0.25)
Ethnicity: Black	0.03 (0.18)	0.02 (0.12)	0.04 (0.20)
Ethnicity: Chinese	0.00 (0.04)	0.00 (0.06)	0.00 (0.04)
Ethnicity: Other	0.05 (0.22)	0.05 (0.21)	0.06 (0.24)
Panel C: School characteristics (7th grade)			
Community school	0.63 (0.48)	0.58 (0.49)	0.71 (0.45)
Religiously affiliated school	0.25 (0.43)	0.28 (0.45)	0.18 (0.38)

Note: The table reports means of the listed variables and standard deviation in parenthesis. The sample only includes pupils in the 50% smallest schools with at most 180 students in the 7th grade cohort, and 135 7th-grade students on average (approx. 4 classes of max 30/35 students). 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 'regular' pupils: approximately 405,000. Number of pupils with at least one subject in top 5% ($\geq$95th percentile of KS2 cohort-specific national distribution): approximately 55,000. Number of pupils with at least one subject in bottom 5% ($\leq$5th percentile of KS2 cohort-specific national distribution): approximately 40,000. 7th grade refers to the first year in secondary school after transition out of primary school. KS3 refers to 9th grade 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
Percentage of new peers for pupils in 7^{th} grade	84.15	31.44	1	99.44
Panel A: Average KS2 percentile treatment (new peers)				
Average peer achievement at KS2 in English	49.78	10.45	1	98
Average peer achievement at KS2 in Math	48.74	9.85	1	100
Average peer achievement at KS2 in Science	48.74	9.93	1	100
Panel B: Top 5% treatment (new peers)				
Percentage, top 5% in English	3.58	3.16	0	19.55
Percentage, top 5% in Maths	3.10	2.65	0	19.87
Percentage, top 5% in Science	3.27	2.82	0	19.86
Panel C: Bottom 5% treatment (new peers)				
Percentage, bottom 5% in English	3.52	3.05	0	19.30
Percentage, bottom 5% in Maths	3.52	2.97	0	19.71
Percentage, bottom 5% in Science	3.54	3.19	0	19.78

Note: The table reports means of the listed variables (except for 'percentage of new peers for pupils in 7^{th} grade', where it reports the median) and standard deviation in parenthesis. The sample only includes pupils in the 50% smallest schools with at most 180 students in the 7^{th} grade cohort, and 135 7^{th}-grade students on average (approx. 4 classes of max 30/35 students). Treatments measured in 7^{th} grade when students start secondary school after transition out of primary school. 'New peers' refer to students in 7^{th} grade in a given cohort that do not come from the same primary school.

Table 3 – Main results: impact of peer quality on KS3 educational attainments and balancing of treatments

	Average peer KS2		*Percentage of bottom 5% pupils*		*Percentage of top 5% pupils*	
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable is:	OLS	Within-pupil	OLS	Within-pupil	OLS	Within-pupil
Panel A: Impact of peer quality on KS3 attainments						
KS3 percentiles, unconditional on KS2; treatments entered separately	0.295	0.014	-1.132	-0.135	0.827	0.031
	(0.015)**	(0.013)	(0.037)**	(0.044)**	(0.044)**	(0.036)
KS3 percentiles, unconditional on KS2; all treatments together	0.117	0.004	-0.803	-0.124	0.464	0.016
	(0.012)**	(0.013)	(0.036)**	(0.044)**	(0.040)**	(0.037)
KS3 percentiles, controlling for KS2 same-subject interacted with subject	0.108	0.002	-0.584	-0.119	0.244	0.038
dummies; all treatments together	(0.011)**	(0.012)	(0.033)**	(0.044)**	(0.036)**	(0.037)
KS3 percentiles, controlling for KS2 same- and cross-subject interacted	0.099	0.002	-0.506	-0.120	0.226	0.043
with subject dummies; all treatments together	(0.010)**	(0.013)	(0.033)**	(0.044)**	(0.035)**	(0.037)
Panel B: 'Placebo-type' test; effect of treatments on KS2 attainments						
KS2 percentiles; all treatments together	0.026	0.012	-0.405	-0.021	0.391	-0.034
	(0.006)**	(0.007)	(0.016)**	(0.017)	(0.017)**	(0.014)*

Note: The table reports regression coefficients and standard errors in parenthesis from regressions of the dependent variable on the treatments. The sample only includes pupils in the 50% smallest schools with at most 180 students in the 7th grade cohort, and 135 7th-grade students on average (approx. 4 classes of max 30/35 students). Standard error clustered at the school level. **: at least 1% significant. The treatment effects in the first row estimated are from two different sets of regressions: one set including average peer achievement at KS2 only (Columns (1) and (2)); one set 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 the 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 listed in Table 1 controlled for in Columns (1), (3) and (5) of Panel A (and 'absorbed' in models with pupil fixed-effects). Number of observations: approximately 1,500,000 (500,000 pupils) in 1190 schools.

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

	Within-pupil estimates		
	Average peer KS2	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
Dependent variable is:	(2)	(4)	(6)
Panel A: Robustness to potential threats to identification			
KS3 percentiles, controlling for KS2: subset of undersubscribed schools	0.011	-0.131	0.061
	(0.015)	(0.048)**	(0.047)
KS3 percentiles, controlling for KS2: student with best subject different from the best subject of new peers (*mixed*)	-0.007	-0.120	0.048
	(0.013)	(0.046)**	(0.039)
KS3 percentiles, controlling for KS2: students with at most one additional school in 'choice set'	-0.004	-0.099	0.031
	(0.012)	(0.046)*	(0.037)
KS3 percentiles, controlling for KS2: students in LEAs with the most restricted choice (less than 12 schools)	0.052	-0.151	0.000
	(0.032)	(0.070)*	(0.068)
KS3 percentiles, controlling for KS2: including school × subject fixed-effects	0.005	-0.086	-0.019
	(0.006)	(0.019)**	(0.015)
Panel B: Results for other samples			
KS3 percentiles, controlling for KS2: students in 33% smallest schools	-0.006	-0.141	0.035
	(0.014)	(0.052)**	(0.045)
KS3 percentiles, controlling for KS2: students in 33% smallest schools, including school × subject fixed-effects	-0.002	-0.115	0.044
	(0.007)	(0.024)**	(0.020)*
KS3 percentiles, controlling for KS2: students in all schools (irrespective of school size)	0.010	-0.090	0.008
	(0.012)	(0.032)**	(0.025)

Note: All specifications include the same controls as in Row (4), Panel A, Table 3. Undersubscribed schools enrol approximately 65% of pupils in the baseline sample. Sample of pupils with different 'best subject' from new peers in school account for approximately 55% of the baseline sample. Sample of pupils with at most one additional school within 'choice set' includes pupils with at most one additional school – besides the one currently attended – within the median home-to-school travel distance from the postcode of residence. Postcodes with at most one additional school represent the median of the school 'choice set' distribution. This sample includes approximately 360,000 pupils. Sample of pupils in LEAs with the most restricted choice includes pupils in LEAs with less than 12 schools. This is the median of the school-per-LEA distribution. This sample includes approximately 140,000 students. Sample of pupils of pupils in the 33% smallest schools includes pupils in schools with less than 158 students in the 7th grade cohort (average cohort size: 120); number of pupils: approximately 297,000. Sample of students in all schools includes pupils in all schools irrespective of their size; number of pupils: approximately 1,350,000. Regressions with school × subject fixed-effects only consider the first cohort (7th grade in 2002) and last cohort (7th grade in 2005). Standard error clustered at the school level, except for regressions with school × subject fixed effects where they are robust. **: at least 1% significant; *: at least 5% significant.

Table 5 – Impact of peer quality on KS3 attainments; breakdown by pupil's gender and ability

	Within-pupil estimates			
	Boys only		Girls only	
	Percentage of bottom 5% pupils	*Percentage of top 5% pupils*	*Percentage of bottom 5% pupils*	*Percentage of top 5% pupils*
Dependent variable is: KS3, controlling for KS2	(1)	(2)	(3)	(4)
Panel A: Pupils of ability pooled (overall effect)				
Overall effect	-0.108	-0.040	-0.125	0.122
	(0.049)*	(0.041)	(0.048)**	(0.044)**
Panel B: Ability blocks defined on original KS2 percentiles				
Effect for	-0.109	-0.025	-0.149	0.148
percentiles below 20	(0.039)**	(0.038)	(0.043)**	(0.042)**
Effect for	-0.126	-0.035	-0.104	0.250
percentiles 20-35	(0.058)*	(0.051)	(0.061)§	(0.057)**
Effect for	-0.100	-0.013	-0.077	0.214
percentiles 35-50	(0.073)	(0.059)	(0.072)	(0.063)**
Effect for	-0.150	-0.046	-0.172	0.096
percentiles 50-65	(0.080)§	(0.061)	(0.070)**	(0.060)
Effect for	-0.141	-0.060	-0.152	0.072
percentiles 65-80	(0.078)§	(0.057)	(0.075)*	(0.061)
Effect for	-0.009	-0.064	-0.078	0.013
percentiles above 80	(0.069)	(0.047)	(0.066)	(0.050)
F-Test: all coeff. jointly equal to zero (p-value)	*0.1088*	*0.8480*	*0.0130*	*0.0003*
F-Test: all coefficients are equal (p-value)	*0.5111*	*0.9416*	*0.4513*	*0.0045*

Note: The table reports regression coefficients and standard errors in parenthesis from regressions of the dependent variable on the three treatments. Treatment effects estimated from one single regression including all three treatments together. Results for average peer quality not reported for space reasons (none of the coefficients was significant at conventional levels). Separate regressions run for boys and girls. Specifications in Panel A are as in Row (4), Panel A, Table 3. Interaction terms in Panel B obtained by interacting the peer quality measures (separately for old and new peers) with a dummy indicating where the pupil ranks in terms of his/her KS2 percentiles *on average across subjects*. 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 the pupil's rank in the ability distribution (and in interaction with subject dummies). All specifications further include subject and subject-by-gender dummies. Number of observations for boys: approx. 745,000 (248,000 pupils) in 1130 schools. Number of observations for girls: approx. 735,000 (245,000 pupils) in 1150 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: treatments separately defined by pupils' gender

	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
Dependent variable is:	(1)	(2)	(3)	(4)
Panel A: Boys only				
KS3 percentiles,	-0.095	-0.134	-0.097	0.016
controlling for KS2	(0.067)	(0.078)§	(0.057)§	(0.063)
Effect size	*0.605*	*0.687*	*0.827*	*0.143*
F-Test: coefficients are equal (p-value)	*0.7205*		*0.1990*	
F-Test: coeffs. jointly equal to zero (p-value)	*0.0643*		*0.2272*	
Panel B: Girls only				
KS3 percentiles,	-0.076	-0.171	-0.011	0.182
controlling for KS2	(0.068)	(0.072)**	(0.062)	(0.061)**
Effect size	*0.470*	*1.072*	*0.088*	*1.978*
F-Test: coefficients are equal (p-value)	*0.3572*		*0.0319*	
F-Test: coeffs. jointly equal to zero (p-value)	*0.0270*		*0.0114*	

Note: The table reports regression coefficients and standard errors in parenthesis from regressions of the dependent variable on the treatments. Treatment effects estimated from one single regression including all 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*) refers to the effect of a one standard deviation change of the within-pupil distribution of peers as a percentage of the standard deviation of the within-pupil distribution of KS3 percentiles. Number of observations for boys: approx. 745,000 (248,000 pupils) in 1130 schools. Number of observations for girls: approx. 735,000 (245,000 pupils) in 1150 schools. Standard error clustered at the school level. **: at least 1% significant; *: at least 5% significant; §: at least 10% significant.

Appendix Table

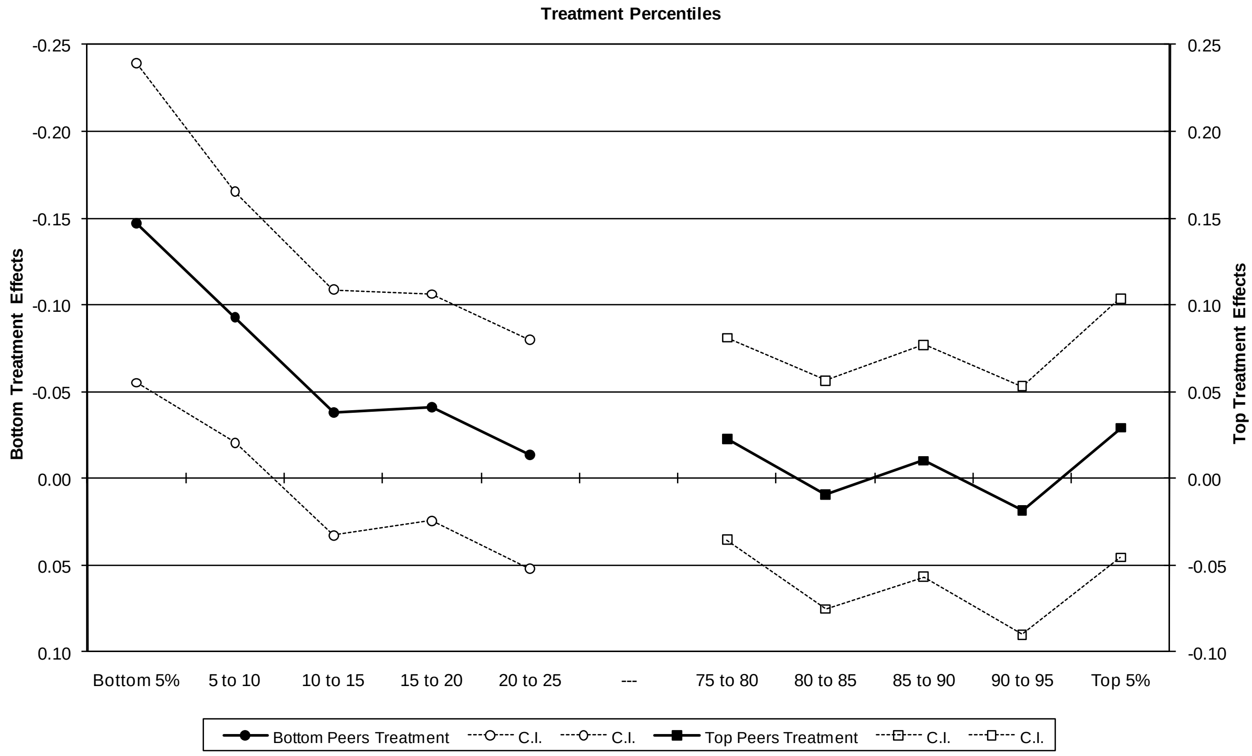
Appendix Table 1 – Within and between variation in pupil test scores and treatment measures

	All students				Sample including boys only				Sample including girls only			
Variable:	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*	*Mean*	*Overall Std.dev.*	*Between Std.dev.*	*Within Std.dev.*
KS2 percentiles	49.61	28.16	25.57	11.80	48.81	28.27	25.57	12.06	50.41	28.02	25.54	11.54
KS3 percentiles	49.94	28.31	26.12	10.94	48.60	28.53	26.23	11.22	51.30	28.03	25.93	10.64
Average peer achievement at KS2	48.76	10.08	9.45	3.50	48.69	10.08	9.47	3.48	48.82	10.08	9.44	3.52
Percentage, top 5%	3.32	2.89	2.56	1.34	3.29	2.87	2.56	1.30	3.44	2.92	2.57	1.38
Percentage, bottom 5%	3.52	3.07	2.89	1.01	3.54	3.08	2.92	1.00	3.51	3.05	2.87	1.02

Note: The sample only includes pupils in the 50% smallest schools with at most 180 students in the 7^{th} grade cohort, and 135 7^{th}-grade students on average (approx. 4 classes of max 30/35 students). Number of observations: approx. 1,500,000 (500,000 pupils), in 1190 schools. Number of observations in samples of boys and girls only: approximately 745,000 (248,000 pupils) in 1130 (boys)/1150 (girls) schools. Peer quality measures refer to new peers only.

Figures

Figure 1 – 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 by 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 and old peer quality; controls for top and bottom old peer quality; subject and subject-by-gender dummies. The sample only includes pupils in the 50% smallest schools with at most 180 students in the 7th grade cohort, and 135 7th-grade students on average (approx. 4 classes of max 30/35 students). Number of observations: approx. 1,500,000 (500,000 pupils) in 1190 schools.