THE UNIVERSITY OF MELBOURNE

FACULTY OF BUSINESS & ECONOMICS

Melbourne Institute Working Paper Series

Working Paper No. 17/16

Achievement Effects from New Peers: Who Matters to Whom?

Duncan McVicar, Julie Moschion and Chris Ryan

MELBOURNE INSTITUTE®
of Applied Economic and Social Research

Achievement Effects from New Peers: Who Matters to Whom?*

Duncan McVicar[†], Julie Moschion[‡§] and Chris Ryan[‡]
[†] Queen's Management School, Queen's University Belfast
[‡] Melbourne Institute of Applied Economic and Social Research, The University of Melbourne
[§] EconomiX, University of Nanterre

Melbourne Institute Working Paper No. 17/16

ISSN 1447-5863 (Online)

ISBN 978-0-73-405212-4

April 2016

* This research was commissioned by the Victorian Department of Education and Training (DET). This paper uses unit record data from the NAPLAN administrative data collection, provided by DET. We thank seminar participants at the Department of Economics, University of Erlangen-Nuremberg; Melbourne Institute Brown Bag, Australia; Department of Economics, University of Tilburg, Netherlands; School of Economics, University of Sydney, Australia; ESAM, Hobart, Australia; and Department of Economics, University of Zurich, Switzerland. The findings and views reported in this paper are those of the authors and should not be attributed to DET or any other branch of the Victorian or Australian Commonwealth government. For correspondence, email <ryan.c@unimelb.edu.au>.

Melbourne Institute of Applied Economic and Social Research
The University of Melbourne
Victoria 3010 Australia
***Telephone* (03) 8344 2100**
***Fax* (03) 8344 2111**
***Email* melb-inst@unimelb.edu.au**
***WWW Address* http://www.melbourneinstitute.com**

Abstract

This paper presents estimates of achievement-related peer effects on school pupils' literacy using data from national test scores, across multiple literacy or language-related measures and student cohorts, for the population of public secondary school pupils in Years 7 and 9 (aged 12/13 and 14/15 years) in the Australian state of Victoria. Identification is achieved via individual fixed effects and by distinguishing between secondary school peers who attended the same primary school as the individual and those who did not. Estimates of peer effects are based on the new peers, whose primary school achievement could not have been affected by the individual. The results provide evidence for the existence of achievement-related peer effects, with small but positive and statistically significant effects from having higher-achieving peers on average and from having a higher proportion of very high achieving peers (in the top 10% of the prior achievement distribution). We do not find a penalty from having 'bad' peers (from the bottom 10% of the prior achievement distribution). Further, it is low achievement individuals who benefit most from having high achievement peers.

JEL classification: I21, I24, J24

Keywords: Peer effects, school achievement, education, tracking

1. Introduction

Peer effects refer to externalities in which the actions or characteristics of a reference group affect an individual's behaviour or outcomes. They have been studied in numerous contexts. This paper examines a specific form of peer effects, related to the effect on a student's achievement of the achievement of her peers, which if the effects are of sufficient magnitude, have critical implications for pupils, parents, schools and policy makers. Effects from peer averages imply that parents can improve their child's expected educational achievement by selecting a school with a higher-ability intake, and reforms introducing greater school choice could widen educational inequalities. Non-linear peer effects have additional efficiency implications, such that schools may be able to improve the average achievement of their pupils, and policy makers the efficiency of the schooling system as a whole, by manipulating the allocation of students across classes or schools. As a result, peer effects in school achievement have attracted an enormous amount of attention in the literature. Establishing the existence and magnitude of peer effects, however, is beset by practical difficulties (Manski, 1993, 2000; Moffitt, 2001, Angrist 2014), a fact that has probably also contributed to the longevity and ubiquity of the literature. Despite the vast number of studies, disagreement as to the nature, magnitude and even the existence of peer effects in school achievement remains.

One such difficulty arises because the individual is a peer of their peers, which means individuals also influence their peers, inducing the potential for reverse causality. One branch of the school achievement peer effects literature has attempted to address this issue by utilising measures of peer quality that pre-date any potential social interactions between the individual and their peers. Two of the more promising studies in this vein are Gibbons and Telhaj (2015) and Lavy *et al*. (2012) (henceforth LEA), both of which exploit the transition between primary and secondary school in England, and the fact that most of the secondary school peers of any individual attended a different primary school. This allows them to use the prior performance of peers when they were in different schools – a measure that is immune from reflection problems – to measure the impact of peers' earlier achievement on current student performance. LEA builds on Gibbons and Telhaj (2015)[1] by also exploiting multiple measures of individual student achievement in various learning domains to remove

[1] Strictly speaking, LEA builds on an earlier version of the Gibbons and Telhaj study released as a working paper in 2008 (Gibbons and Telhaj, 2008).

any fixed student-specific effect on achievement, which mitigates confounding effects from sorting into schools. LEA finds no evidence of an average peer effect on individual achievement; Gibbons and Telhaj (2015) find evidence of a small positive effect, with a one standard deviation increase in average peer ability leading to a 0.02 standard deviation increase in own achievement.

The research design adopted in this paper also exploits this transition, when students encounter new classmates at the start of secondary school who did not attend the same primary school. We regress the secondary school achievement of individuals on the primary school achievement of their secondary school new peers, for whom there is no reverse causality. Our approach also makes use of multiple measures of individual achievement, in our case a set of language or literacy-related measures of achievement. This enables us to use an individual fixed effects approach, similar to that of LEA, to account for common individual effects on achievement across the four language-related learning domains of reading, writing, grammar and spelling. Because we have multiple cohorts of students, we also use school-subject fixed effects to wash out time-invariant correlated effects that might differ across subjects. In other words we estimate peer effects by exploiting within-individual effects from variation in the performance of new peers across the literacy domains when they were in primary school. Our main result on average peer effects turns out to be very different to that of LEA.

In addition to estimating average peer effects, LEA also assesses the impact on an individual of having a high proportion of peers who were in either the top or bottom 5% of the achievement distribution, while also allowing these effects to differ between genders and across the distribution of individual ability (that is, they allow for non-linear and heterogeneous peer effects). This allows them to test empirical support in English secondary schools for a wide range of models including the 'bad apple' (disruptive students harm everyone), the 'shining light' (excellent students provide a great example for all), the 'invidious comparison' (outcomes are harmed by the presence of better achieving peers), and ultimately to assess the possible benefits of tracking, at least at the school grade level (see Sacerdote, 2011). Here they do find evidence of (bad-apple type) peer effects. We too follow this approach, again with contrasting results.

Our main points of departure from LEA are that we use entirely new data for a different country (Australia rather than England); that we focus on four language-related learning domains rather than the LEA subject-set of English, mathematics and science; and that we focus mainly on outcomes for a younger age group than LEA (the first year of secondary school in Victoria, aged 12/13 years, rather than the third year of secondary school in England, aged 13/14 years. These differences with LEA turn out to be crucial: in many respects our results are the polar opposites of those in LEA. A potential explanation for this stark contrast is that in the LEA study grade-level measures of peer ability are a poor proxy for class-level peer ability – with classes being the level at which achievement peer effects are most likely to exist – in the presence of subject-specific ability streaming (tracking) within grades. In our case, because tracking is much less common in language-related subjects than in mathematics and science, much less common in the first year of secondary school than in later years, and less prevalent more generally in Australian public schools than in English public schools, our results should be comparatively free from this source of attenuation bias.[2] So where LEA find zero peer effects at the average and from the proportion of high-achieving peers in the grade, we find positive and statistically significant peer effects in both cases.

Specifically, we estimate peer effects in school achievement using administrative data on the test scores of school pupils at public schools in the Australian state of Victoria. Our data come from the National Assessment Program – Literacy and Numeracy (NAPLAN) conducted across Australia, which provides test scores for five achievement domains – numeracy, reading, spelling, grammar, and writing – for all pupils in grades 3, 5, 7, and 9. The national testing system which generates these data was introduced across Australia in 2008, with tests taking place each year in the specific grades, so we are able to exploit data from 2008 to 2013. This provides us with four cohorts of Year 5 (primary school) students subsequently observed in Year 7 (in secondary school), and two cohorts with students also observed in Year 9.

[2] For example, LEA quote figures (p379) that ability tracking takes place within 46%, 59% and 80% of schools at Year 9 for English, science and mathematics, respectively. Thomson et al. (2003, p105) report comparable Australian figures of 58% for maths and <20% for science in Year 8, which are likely to be upper bounds on the extent of ability tracking in Year 7 given tracking increases with age and tracking is more common in non-government schools within Australia. Forgasz (2010) reports survey-based proportions of 37% (70%) of public schools in Victoria ability track in mathematics in Year 7 (Year 9).

The remainder of this paper is set out as follows. The next section briefly reviews relevant studies in the school achievement peer effects literature, and in particular LEA, to which this paper most closely speaks. Our data are described in more detail in section 3, along with further discussion of the approach to estimation. The main results are discussed in section 4, with extensions in section 5 and concluding remarks in section 6.

2. Literature

Recent studies on peer effects in school achievement have looked for both average peer effects and peer effects from and to different points in the ability distribution, using a variety of strategies to overcome the identification problems associated with reflection and endogenous sorting. One strand of this literature exploits random or quasi-random assignment to new peer groups stemming from programs like Metco in the US (Angrist and Lang, 2004) or the Extra Teacher Program in Kenya (Duflo et al., 2011). These studies make credible claims to estimating causal peer effects (or more precisely the absence of causal peer effects) that are internally valid. But a dearth of suitable similar experiments in other contexts means we cannot yet draw general conclusions from this strand of literature alone. A second (non-experimental) group of studies regresses own achievement on lagged peer achievement under the assumption that an individual's current achievement cannot impact her peers' prior achievement (e.g. Hanushek et al., 2003; Lefgren, 2004; Vigdor and Nechyba, 2007; Atkinson et al., 2008). Manski (1993) questions the extent to which this approach truly solves the reflection problem on the grounds that peer scores are likely to be persistent. Additional identification steps – typically school fixed effects and individual controls – are also required to try to wash out sorting effects, arguably with mixed success. Angrist (2014) also critiques this approach for being susceptible to 'mechanical' biases due, among other things, to measurement error and negative intra-group correlation between own test score and the leave-out-mean test score of peers. A third group of studies exploits demographic variation in peer group composition to instrument for peer achievement (e.g. Goux and Maurin, 2007), with the extent to which the resulting estimates can be interpreted as causal peer effects hanging, as ever, on the plausibility of the exclusion restrictions. Sacerdote (2011) provides a review, summing up the body of evidence from this literature as "mixed".

The current paper builds on a fourth and as yet small group of studies – Gibbons and Telhaj (2015) and LEA – which arguably falls somewhere between the first and second literature strands. Like the second group of studies described above, these studies also regress own achievement on lagged (school grade) peers' achievement, with a number of additional identification steps to address sorting and other correlated effects. Like the first set of studies described above, however, they focus on the impact of being grouped with *new* peers, in this case as a result of the transition between primary and secondary school in England. These studies are immune to reflection problems because the prior scores (in primary school) of new peers (in secondary school) cannot have been impacted by the individual. Although both studies predate Angrist (2014), this research design mitigates problems associated with "Angrist mechanics" such as the negative intra-group correlation. Further, LEA make a convincing case that they wash out correlation due to endogenous sorting by basing their conclusions only on within-individual variation in educational achievement across different learning domains (English, mathematics and science).[3] LEA finds no evidence of average peer effects but finds that high achieving peers impact positively on low achieving girls (but not boys), and that both genders across the ability distribution suffer from having low achieving peers (bad apples). These effects, although statistically significant, are small in magnitude. Gibbons and Telhaj (2015) do find evidence of average peer effects, which again are small in magnitude.

These studies make a serious contribution to the peer effects literature. We expect numerous studies to follow their lead, contributing to what may become a substantial body of evidence on school achievement peer effects in its own right. Indeed, Mendolia et al. (2016) is another very recent contribution to this strand of the literature, also exploiting data on the primary to secondary transition in English schools, although it takes a slightly different approach in using primary school test scores of peers of peers to instrument for the secondary school test scores of peers. As it stands, however, there is a question mark over the extent to which the conclusions of LEA, and to some extent the other papers in this emerging strand of the literature, will generalise. The particular context for the LEA study - England, at a time of major school choice reforms, and with a high degree of within-grade subject-specific tracking in the relevant age group – may be atypical of public school systems elsewhere and at other times. Most importantly, the high degree of tracking in the schools and year group studied by LEA means their estimates of peer effects, relying as they do on grade-level proxies for what

[3] Although LEA were the first to use this within-individual cross-domain approach to examine peer effects in school achievement, it has been used in various studies of other education-related questions (see, for example, Dee, 2005; Altinok and Kingdon, 2012; Schwerdt and Wuppermann, 2011; Lavy, 2010).

are predominantly class-level effects, are likely downwards biased.[4] In applying the LEA research design across literacy-related learning domains in Australian secondary schools, and in the first year of secondary education during which tracking is less common than in later grades, we hope to make a significant additional early contribution to this body of knowledge.[5]

3. Data and Identification

The national testing system, NAPLAN, was introduced across Australia in 2008, with tests taking place each year in grades 3, 5, 7 and 9. We have test score data from 2008 to 2013 for the population of students in the relevant grades in public schools. This provides us with four cohorts of Year 5 (primary school) students who are subsequently observed in Year 7 (in secondary school) from 2010 to 2013. In addition, we have two cohorts with students observed in Year 9 in 2012 or 2013. For both the Year 7 and Year 9 cohorts, secondary school peer groups are defined using the Year 7 data, and we distinguish between Year 7 peers who attended the same primary school (old peers) and those who attended a different primary school (new peers). For both groups, we use their Year 5 achievement outcomes as measures of new and old peer group quality.[6] We focus primarily on the results from the analysis of Year 7 literacy, though we also report those for Year 9.

NAPLAN scores across all grades are measured on a scale from 0 to 1000, and are designed to measure absolute (rather than relative) competence in the subject concerned.[7] Following LEA, we convert these scores within individual learning domains, by year and grade, but across all schools, to percentile ranks. There is some 'lumpiness' in the scores, since they are based on Rasch modelling, so there are as many distinct scores in each subject as there were test items. In the analysis, all individuals with the same subject score are given the same rank.

[4] LEA states the following (p378):"…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." Gibbons and Telhaj (2015) and Mendolia et al. (2016) similarly use grade-level data with widespread within-grade ability tracking.

[5] There are also few credible studies of peer effects in school achievement using Australian data, so policy makers and school principals in Australia face important questions about allocation of pupils between and within schools with very little evidence on which to base their decisions.

[6] Some schools, particularly in remote areas, encompass both primary and secondary school stages, so most secondary students in such schools only have old peers. These schools are excluded from the analysis. Only 1.2% of students are dropped of our sample because they have no new peers.

[7] Average NAPLAN scores in a particular subject area are therefore higher in higher school grades. Note that just under 7% and 9% of students at Years 7 and 9 respectively, were not entered for, or otherwise missed, NAPLAN tests in 2010. Less than 5% of students did not sit the tests in Years 3 and 5 in 2010.

This approach avoids the problem that scores across different learning domains in NAPLAN are not intended to be directly comparable. The inclusion of separate domain or subject-level effects in the regression equation, again following LEA, is also intended to deal with this issue.

One feature of the analysis in LEA – their attempt to address the ability tracking problem – is that they restrict their data to the smallest 50% of secondary schools in England (they assess how sensitive their results are to this restriction by re-estimating with the smallest third of schools). Note, however, that even in the latter case the average grade size is still 120 students, so these are likely to be schools with four classes per grade on average, and are therefore still likely to feature ability tracking. Nevertheless, they retain over a million student observations. We do not have so many observations, and neither do Australian schools track Year 7 students in literacy-related subjects to the same degree as English schools track Year 9 students, particularly in mathematics and science, so we do not make the same restriction. We do, however, exclude students from four public schools with selective entrance policies. Since these operate from Year 9, this exclusion does not affect the results for Year 7 students in any way. We also examine whether estimated peer effects vary by school size and by the proportion of secondary school grade-mates who are new peers, itself correlated with school size.

The process by which peer groups are constructed is demanding. Scores for all identified students in all four subjects are required in the current year, as well as in the past. Since many students miss one test in a particular year, limiting the sample to only those with complete responses and allowing the peer effects to reflect only the achievement percentiles of those with complete data reduces the sample. Therefore, we construct two data sets – one where only individuals with complete information were included (the *restricted* sample) and another where those with partial information were also included (the *unrestricted* sample), which increases the number of observations by about 10,000 for both the Year 7 and Year 9 estimates. The results reported are those for the *unrestricted* sample, though those for the *restricted* sample are similar and available from the authors.

In addition to test score data, we can identify each pupil's school and grade (but not class), which we use to construct peer test scores. We also have information on gender, date of birth, an indicator of Indigenous status, an indicator for whether English is the main language spoken at home, and information on mothers' and fathers' occupations and education levels. These are included as controls at the individual level in the OLS models but not the individual fixed effect models. We also use this information to examine evidence for heterogeneous peer effects.

Table 1 gives summary statistics for Year 7 students, for all students, 'regular' students (those whose scores lie between the 10^{th} and 90^{th} percentiles), as well as for those with at least one subject where they were in the top 10 percent of students and for those with at least one subject where they were in the bottom 10 percent of students. Unsurprisingly, the typical ranking in any subject is always close to 50. More surprisingly, the average rank of students with at least one score above the 90^{th} percentile is 10 percentile points lower than 90, suggesting the average rank of other scores is often a rank within the 70s, well down on their maximum score. As expected, higher performing students have parents with higher levels of education and who work in management and professional or associate professional occupations. Males and those from Indigenous backgrounds are more likely to have achievement in at least one subject in the bottom 10 percent of students.

The size of the average Year 7 peer group is around 175 students. From Table 2, about 85 percent of a student's peers in Year 7 were new peers. Their average percentile rank was a little below average, their share being in the top 10 percent a little lower than 10 percent, and their share in the bottom 10 percent a little higher.[8] Table 3 contains an analysis of the variance in the new peer percentile rank variables. Most of the variance arises from between peer group variation rather than within peer groups, suggesting, as in LEA in the English case, that sorting between schools is important.

One feature of our data is that they include only students in public schools. Yet the transition from primary to secondary school is precisely the time when many students in Australia exit the public system to attend private schools. Although these lost peers do not have any internal validity implications for our analysis, they may limit the inferences we can make beyond the specific set of individuals observed in public secondary schools in the relevant years. These private schools remain a key concern of policy makers in Australia, in part because they

[8] The departures from 10% reflect the movement of higher achieving students from public to private schools in the transition from primary to secondary education common in Australia. Hence, more that 10% of the remaining students come from the original bottom decile, fewer than 10% from the original top decile.

typically receive substantial public subsidies, but also because they draw disproportionately from the upper end of the ability distribution, which may have both equity and efficiency implications in the presence of peer effects. We return briefly to this coincident transition between public primary and private secondary schooling in the conclusion.

Our starting point for estimation allows for individual, subject and school-subject fixed effects, denoted in equation (1) below by α_i, β_q and η_{qs} respectively. We estimate these effects separately for grades 7 and 9. Following LEA, we allow the subject effects to vary by gender, though note that we are considering subjects only in the broad language domain. In addition, we anticipate individual *i*'s achievement *A* in subject *q* in grade *g* will be influenced by peer average effects measured by performance in an earlier grade (*g* = *c*), denoted by $\overline{A}_{qcst}$, as well as by the proportion of high (top 10%) and low (bottom 10%) achieving peers they have, denoted by $\bar{P}^{h}_{qcst}$ and $\bar{P}^{l}_{qcst}$, respectively. Rather than entering these peer effects aggregated to the whole grade level, however, we distinguish between the effects of 'new' and 'old' peers via $\bar{A}^{n}_{qcst}$ and $\bar{A}^{o}_{qcst}$, where 'new' peers are those in grade *g* at school *s* with individual *i* who did not attend the same school as individual *i* in grade *c,* while the 'old' peers attended the same school as individual *i* in grades *c* and *g*, We make the same distinction for the variables denoting high achieving and low achieving peers. Further, an individual *i*'s achievement in subject *q* in grade *g* will also be influenced by her prior achievement in grade *g* = *c* in that subject, A_{iqcst}, along with the cross-subjects (*q* = *d*, *e, f*) A_{idcst}, A_{iecst} and A_{ifcst}(with subject specific coefficients). This gives us equation (1) as:

$$(1) \qquad A_{iqgst} = \alpha_i + \beta_q + \beta_q \times gender + \eta_{qs} + \delta_1 \bar{A}^{n}_{qcst} + \delta_2 \bar{A}^{o}_{qcst} + \theta_1 \bar{P}^{nh}_{qcst} + \theta_2 \bar{P}^{nl}_{qcst} + \theta_3 \bar{P}^{oh}_{qcst} + \theta_4 \bar{P}^{ol}_{qcst} + \lambda_{1q} A_{iqcst} + \lambda_{2q} A_{idcst} + \lambda_{3q} A_{iecst} + \lambda_{4q} A_{ifcst} + \varepsilon_{iqgst}$$

where δ_1, δ_2, θ_1, θ_2, θ_3, θ_4, λ_1, λ_2, λ_3 and λ_4 are parameters to be estimated and ε_{iqgst} is a random error term.

The main coefficients of interest are δ_1, θ_1,and θ_2, the parameters on the prior test scores of new peers and on the proportion of new peers who are high or low ability. We do not interpret estimated associations with old peer test scores as peer effects, although they may be partly driven by peer effects. Rather, old peers are included as controls for shared primary school*cohort*subject effects that might persist into secondary school. The inclusion of school*subject dummies in equation (1) is to control for possible time-invariant subject-specific sorting into schools. While such effects are less likely in our case than in LEA, given that our learning domains are all literacy-related, including these effects nevertheless slightly lowers the magnitude of the estimated peer effects.

Although we estimate equation (1) using different data, this specification is identical to the preferred specification of LEA, other than in the respect that our high/low achieving peer variables are defined at 10% thresholds rather than 5% thresholds and that we include school by subject fixed effects instead of their school by year fixed effects.[9] The identification arguments of LEA therefore apply similarly here: the various fixed effects wash out potentially confounding effects from sorting and other correlated effects, and we are immune from reflection problems and common shocks (e.g. teacher effects) because we focus only on the impact of prior test scores of new peers and in a single subject area. Given our data, there is no perfect placebo test to examine whether any biases still remain in our most complete specification. But we can get close to such a test by replacing the individual's current (Year 7) test score with her Year 5 test score, predating the transition to secondary school. Peer effects should not be possible in this scenario because the individual is yet to meet her new peers. But any sorting effects would still show up as correlation between own and new peer scores in Year 5. The one weakness of this test is that we cannot include own prior score as a control, so it is a placebo test on not quite our preferred specification. LEA estimates a similar placebo test.

LEA predates, and therefore does not directly address, the estimation issues raised by Angrist (2014). Their (and our) research design, however, mitigates the threat to placing a causal interpretation on the resulting estimates from the "Angrist mechanics". The leave-out mean is not used, so one source of mechanical bias – from the negative correlation between own score and peer score within group – is avoided. Systematic rather than random variation in group composition means weak instrument bias is also unlikely to be an issue.[10] Peer groups are also not symmetric in the sense that they vary within the secondary school grade for those

[9] The 10% threshold is the one where the estimated top/bottom peer effects are largest. We discuss sensitivity to the particular choice of threshold in Section 4.2. We do not include school by year fixed effects because they did not change the estimates that did not include school by subject fixed effects.

[10] The within-individual design and the suite of other included fixed effects mean we do not pay an endogenous sorting price for this non-random group composition.

from different feeder schools. To support these claims we carry out a further placebo test which keeps the real peers of individuals but randomly mixes average peer scores across the different subject domains for each individual.

Like LEA, what we estimate here are 'net' peer effects, i.e. net of ability spillovers between subjects. LEA warns that if such spillovers were strong enough they could mask subject-specific spillovers so that we would find zero peer effects. Because our domains are all literacy-related, it may be that cross-subject spillovers are more of an issue for us than for LEA, so we are in effect setting a high bar for peer effects here.

4. Main Results and Discussion

4.1 Homogenous Peer Effects in Year 7

Table 4 contains parameter estimates on the new peers variables included in equation (1) for Year 7 students. The first three columns show the parameters on the average, bottom 10% and top 10% new peer effects estimated by least squares with a set of individual controls but no individual fixed effects. These are included for completeness. We focus our discussion on the next set of three columns which show the parameters when the equation is estimated with student fixed effects - the 'within' student estimates. Each subsequent row in Table 4 corresponds to a version of (1) with an increasing degree of control. As we will see, estimated peer effects tend to fall in absolute value as we include progressively tighter controls.

The first row shows estimated associations between own Year 7 score and prior scores of new peers when each of the peer variables are entered singly in equation (1), but without controlling for old peers or any individual prior achievement scores. The estimates point to substantial and positive within-individual association between own score and the average prior score of new peers, as well as significant associations with the proportion of new peers in the bottom 10% of the ability distribution (negative) or with the proportion of new peers in the top 10% of the ability distribution (positive). The effects are of the order that a one point increase in the average percentile rank of new peers is associated with a 0.4 unit rank increase in the subject's percentile rank, and a one percentage point increase in the proportion of top and bottom students is associated with a 0.3 and -0.2 change in the percentile rank of the individual, respectively.

The second set of results show the parameter estimates when the three new peer variables are entered simultaneously in equation (1), but still without controlling for old peers or individual prior achievement. Inclusion of the proportions at the top and bottom of the achievement distribution has only a small effect on the average effect parameter, but including the average substantially reduces the parameter on the effect of the proportion of top and bottom achievers. From this point on, none of the estimates point to there being any detrimental impact from having a higher share of peers in the bottom 10% of the achievement distribution.

The third set of results show the parameter estimates when the three peer variables are entered in equation (1), this time controlling for the old-peer equivalents, but still without any individual prior achievement scores. As in LEA the old peer variables are intended to control for shared primary school*cohort*subject effects that might persist into secondary school. As in LEA, the estimated parameters on the new peer variables change little with the inclusion of the old peer variables.[11]

The fourth set of results show the parameter estimates when we add controls for same-subject own prior score interacted with subject dummy variables. This is intended to address sorting into schools that might be subject specific, in addition to the individual fixed effect which washes out non-subject specific sorting. The interaction with subject dummies allows own prior score to impact differently in different subjects. Again, both remaining associations between own (Year 7) score and the scores of new peers fall in absolute value, but remain statistically significant.

The fifth set of results builds on the fourth by including cross-subject own prior scores interacted with subject dummy variables. Again we follow LEA in taking this step, which allows for cross-subject spillovers from own prior scores that differ by subject. This is the preferred specification of LEA, and they interpret the resulting estimates causally, although they subsequently conduct further sensitivity analysis to support this interpretation. In our case inclusion of these additional controls lowers the estimated magnitudes of the effects. In these estimates, the effects are of the order that an increase in the average percentile rank of new peers by 1 unit rank is associated with an individual's percentile rank that is around 0.1

[11] Note that, while not reported, the estimates on the old peers variables are always smaller in absolute values than the new peer values reported in Table 4.

of a unit rank higher. A one percentage point increase in the proportion of top students is associated with a 0.03 higher percentile rank in the subject, substantially smaller effects than those reported in the simplest specification.

There is clearly a major departure in our results compared to the equivalent specification in LEA. In the aggregate, they found no average peer effect, with a coefficient very close to zero and nowhere near statistical significance. They also found no statistically significant effect from top peers, although their point estimate is actually similar in magnitude to ours. They did, however, find a negative effect from low-achieving peers, with a one percentage point increase in the proportion of low achieving peers leading to a fall in own performance by 0.12 of a unit rank. In aggregate, our results for Year 7 in Australia are almost the complete reverse of this: there appears to be a small but highly statistically significant average peer effect, a small but statistically significant positive effect from top peers, and no negative effect from bottom peers.

What drives this stark contrast? There are a number of potential explanations. Some, however, can be dismissed by appealing to the literature and/or the sensitivity analysis set out in Section 4.2. First, we study different learning domains to those studied by LEA, and if peer effects vary by subject then we might perhaps expect to find some difference in results. On its own, however, we would expect this to lead to stronger peer effects in LEA than in our case. Bettinger (2012), for example, argues that mathematics test scores may be more elastic to a range of school-based interventions than reading scores because mathematics learning takes place disproportionately in schools whereas reading takes place disproportionately outside school. Our own sensitivity analysis (see section 4.2) also suggests estimated average and top peer effects are either unchanged or larger when we add numeracy to the set of included learning domains.[12] Related to this, differences in cross-subject spillovers between the sets of subjects studied in the two papers may also play a role; at least to the extent that they are not controlled by own prior cross-subject scores. But because our subjects are more similar than the LEA subjects this also seems more likely to act against us finding peer effects in the fixed effects model compared to LEA. The set of subjects studied by LEA may also be more susceptible to subject-specific correlated effects, e.g. from subject-specific sorting into schools, to the extent that any such effects remain after controlling for own prior

[12] Also see Gibbons and Telhaj (2015) on this point.

score. This too would make estimates of positive peer effects comparatively more likely for LEA. Both our own and LEA's sensitivity analysis, however, including in particular models where school by subject dummies are added, suggests that this is not a key issue (see below and section 4.2). Nevertheless, there is one contrast relating to the different subject sets in our paper and LEA that is more difficult to dismiss, and that's the greater degree of ability tracking within schools in mathematics and science than in language-related subjects. We would expect less downwards bias in our estimates than in those of LEA as a result.

Second, the two papers adopt different thresholds for the low/high achieving peers variables. Both we and LEA show the robustness of these estimates to different thresholds, however, and no alternative threshold overturns the contrast in results. Having said that, when we use the 5% threshold, the 'shining light' effect does fall slightly in magnitude and drops below the 95% level of statistical significance, although this has no bearing on the contrasting results for average peer effects (see Section 4.2).

Third, we study outcomes for a younger age group than LEA – 12/13 year olds in their first year of secondary school compared to 13/14 year olds in their third year. In both England and Australia the extent of ability tracking within grades tends to increase grade by grade (LEA; Thomson et al., 2003; Forgasz, 2010), so by focussing on a younger age group our grade-level peers are likely to be better proxies for class-level peers than is the case for older age groups. LEA's fix for this – restricting their attention to smaller schools which should feature less tracking – seems unlikely to address the problem. Again we would expect less downwards bias in our estimates than in those of LEA. Our own estimates for Year 9 in Victoria show the effect from high achieving peers becomes statistically insignificant, although the average peer effect remains robust.

Fourth, given grade and subject, public schools in Australia ability track to a lesser extent than in England, so even if we were to restrict our attention to Year 9 and small schools, we would still not be comparing like with like in terms of tracking, and again we might expect less downwards bias in our estimates than in those of LEA. (In fact, when we restrict out attention to Year 9 and small schools, our estimates are closer to those of LEA – see Section 5.2.) Other institutional and cultural differences, including the potential for differences in teaching approaches, may also play a role in an uncertain direction.

The sixth and final set of results in Table 4 builds on the fifth by including school by subject dummies to wash out any remaining time-invariant subject-school unobservables, e.g. subject-specific sorting reflecting school specialism in a particular subject. This is our preferred specification, since it offers a further degree of control over and above specification 5 above, and these results are the first for which we tentatively claim a causal interpretation. Again there is no impact from low-achieving peers in the bottom 10% of the achievement distribution (the coefficient is negative but very small and nowhere near statistical significance). There is also little change in the estimated impact from high-achieving peers. The effect from the peer average decreases by a third. A one percentage point increase in the proportion of students in the top 10% of the achievement distribution increases own score by 0.04 of a percentile rank. A one point increase in the average percentile rank of new peers leads to a 0.08 increase in the percentile rank of the individual in Year 7. LEA also estimates a version of their model with school*subject dummies, and although precise magnitudes change[13] their overall conclusions do not.

4.2 Robustness

The outcomes of a number of robustness tests, in each case of the sixth specification of the model with the highest degree of control, are described in this section. First note that using the restricted sample (only those individuals for whom we observe test scores in all subjects at both Year 5 and Year 7) in place of the unrestricted sample leads to almost identical estimates. Second, note that limiting the analysis to 'regular' students (those without test scores in either of the top or bottom 10% of the distribution) also has little impact on the results, as in the case of LEA. Third, we report on a version of (1) where we add numeracy to the set of student achievement measures studied.[14] Fourth, we use raw test scores in place of percentiles. Finally, we explore robustness of the estimated impacts of high/low achieving peers to varying the thresholds at either end of the distribution.

[13] The magnitude of their estimated low-achieving peers effect falls by a factor of around one third, which is similar to the fall in the magnitude of our average peer effect when adding school*subject dummies. Their positive but non-significant estimated effect of high-achieving peers switches sign but remains insignificant. There is little evidence here, then, that the LEA study is more susceptible to subject-specific correlated effects than the current study; at least in terms of time-invariant correlated effects.

[14] For conciseness we do not include tables of results for these sensitivity analyses. They are available from the authors on request.

The addition of numeracy to the set of student tests considered – moving us away from only considering literacy-related learning domains and closer to the subject mix of LEA – does not change the estimated average peer effect or the zero effect from low-achieving peers, but does lead to a larger estimated impact of the top peer effect (it increases to 0.057 from 0.038). This may in part reflect the kind of arguments put forward by Bettinger (2012) for differences in behavioural responsiveness between reading and mathematics. But at least it suggests that differences in subject mix between LEA and the current study are not in themselves behind the contrasting results. Note that the difference between the 5^{th} and 6^{th} specifications – with and without school*subject dummies – with numeracy is a little larger than in the literacy-only case, consistent with there being more scope for school specialization when considering the wider set of subjects.

We have followed LEA in converting test scores to percentiles, which are more comparable across grades and subjects than the raw scores. But the inclusion of subject dummies in (1) and the grade-level focus should mean that our estimates are not qualitatively sensitive to this conversion. This is indeed the case: estimating (1) using raw scores suggests a small but positive and statistically significant average peer effect, a positive and statistically significant impact from high-achieving peers in the grade, and zero impact from low-achieving peers.

Finally, we explore sensitivity to setting the threshold at 10% for the high-achieving and low-achieving peer variables. Figure 1 shows how the estimated low and high-achieving peer effects vary as we vary the threshold from top/bottom 5% through to top/bottom 25%. There is no threshold for which the bottom peer effect is negative and statistically significant, and the biggest negative impact is where the threshold is set at 10%. In contrast, the statistical significance (but not the sign) of the high-achieving peer effect is sensitive to threshold, again with the biggest positive impact corresponding to the 10% threshold. It is the high-achieving peer effect, for which estimates in both the current paper and LEA appear most sensitive, where the contrast with LEA is least clear.

4.3. Placebo Tests

We tentatively interpret the estimates for the sixth specification in Table 4 as causal. In order to be more confident in this interpretation we want to be sure that we find no 'evidence' of peer effects in scenarios where peer effects are not possible, by design. The biggest remaining concern in terms of how confidently we can interpret our estimates as peer effects are whether there might be any remaining subject-specific sorting that is not washed out by own prior scores and school*subject dummies. We conduct two placebo tests to examine these threats to our causal interpretation. Results are presented in Table 5.

The first test follows LEA in replacing own current score with own prior score on the left hand side of (1). (Own prior score therefore obviously gets dropped from the right hand side of (1).) Peer effects should be impossible in this scenario because the individual is yet to meet her new peers. But any remaining sorting effects would still show up as correlation between own and new peer scores in Year 5. Encouragingly, both the average peer effect and the high-achieving peer effect disappear in this case. The low-achieving peer effect becomes statistically significant at the 90% level, although it is small in magnitude. Nevertheless the test suggests broad support for our interpretation of the relevant Table 4 estimates as capturing causal peer effects. Were we to adjust our Table 4 coefficients by subtracting the relevant Table 5 coefficients, only the estimated impact of high-achieving peers would be qualitatively sensitive. Both the zero peer effect from low-achieving peers and the positive average peer effect would remain unchanged.

The second test randomly mixes the achievement scores of an individual's new peers across the different subject domains. Since within-individual variation in average peer scores across subjects is the main source of variation we exploit in estimation of peer effects in this paper, it is critical to see that these disappear if the peer scores are randomised. Peer effects should be ruled out in this scenario, even if there are cross-subject spillovers, by this randomisation. We find that the estimated effects are all small in magnitude and not statistically different from zero.

5. Extensions

5.1 Heterogeneity in the Year 7 Peer Effects

Studies have examined evidence for heterogeneous peer effects by characteristics like gender, race and ability in addition to homogenous peer effects. Table 6 contains parameter estimates on the new peers variables included in (specification 6 of) equation (1) for males and females and across the distribution of prior school achievement. We are interested in whether and how the peer effects might be different for these groups. LEA, for example, find evidence of some heterogeneity in estimated peer effects along these dimensions. Specifically, they find that girls but not boys benefit from interacting with high-achieving peers, particularly girls in the bottom half of the prior achievement distribution. The effect of the bottom 5% and the zero average peer effect, however, are common to boys and girls and are quite stable across the ability distribution, with the possible exception of the top 20% of the distribution for whom low-achieving peers have less impact (statistically insignificant for both genders). In our case, Table 6 panel A shows that the significant average new peer effects are common to boys and girls, though the magnitude of the average effect is marginally larger for females. The positive effect from the proportion of high-achieving students is predominantly a female phenomenon. The estimated parameter is smaller and not significantly different from zero for males. Note the consistency with LEA in this respect; the difference is that their estimate becomes statistically insignificant when boys and girls are pooled and ours does not. The zero impact from low-achieving peers is common to both genders.

The new peer effects across the distribution of Year 5 achievement are broadly similar for males and females. The average new peer effect is larger for those in the top half of the achievement distribution for both genders. For both males and females, the high-achieving peer effect is larger for lower performing students than for those in the top half (boys) or top third (girls) of the distribution of prior achievement. So our evidence is consistent with 'shining-light' and inconsistent with 'invidious-comparison' peer effects across the ability distribution. (Note a high degree of ability tracking within grades would limit the interaction between high and low achieving peers, making such peer effects less likely.) The impact from low-achieving peers is almost everywhere statistically insignificant for both genders, but does tend to change from small negative to small positive as we move up the prior achievement distribution.[15]

LEA limit their sample to smaller schools in order to reduce the extent of within-school-grade tracking. We do not make this restriction in our main estimates because we do not believe it is an effective fix for tracking, we have less tracking in any case, and for sample size reasons. Nevertheless, there may be other differences between small and large schools that, if peer effects are heterogeneous, may be relevant for drawing comparisons between the results in

[15] We do not read much into the marginally significant positive effect of low-achieving peers for high achieving boys, and the negative effect of low-achieving peers for low achieving girls.

LEA and those in the current paper. Table 7 reports estimated new peer effects by school size (above and below the median). Limiting the sample to students from small schools does have an impact on the estimated average peer effects, with the relevant coefficient falling from 0.081 in Table 4 to 0.039 in Table 7 (significant only at the 10% level) when only students in the smallest 50% of schools are used, and to 0.030 (and not significant even at the 10% level) when only students in the smallest 33% of schools are used. Were this capturing less tracking within grades in smaller schools – as LEA maintain for their sample – we would expect estimated peer effects to increase in magnitude. The coefficients on the number of low-achieving and high-achieving peers change little, although slight increases in the standard error mean the high-achieving peer coefficient is only statistically significant at the 10% level when the sample is restricted to smaller schools.

Table 7 also contains estimates where the proportion of new peers in the grade is split into quintiles. Larger schools tend to have more primary school feeders and therefore a higher proportion of new peers in Year 7. For both small and large schools the average effects of new peers tend to increase with the proportion of new peers one has, although with a statistically insignificant tailing off for the top quintile. There is no clear pattern by proportion of new peers in the estimated impacts of high-achieving peers, and higher standard errors tend to make these estimates statistically insignificant. The same goes for the impact of low-achieving peers in large schools: there is no pattern and it is everywhere statistically insignificant. There appears to be more of a pattern in the estimated impact of low-achieving peers in small schools, however, which increases in absolute magnitude with the proportion of new peers, becoming statistically significant and negative for the highest quintile. Our conjecture is that this, and the heterogeneity in the average peer effect, may reflect friendship links within the grade, whereby individuals in small schools with fewer new peers may disproportionately maintain friendships with peers from the same primary school, reducing their exposure to new peers.

5.2 Peer effects for Year 9

Table 8 contains parameter estimates on the new peers variables included in equation (1) for Year 9 students. Own and peer prior scores are measured in Year 5, as before. Our motivation for estimating Year 9 peer effects in addition to the Year 7 peer effects is twofold. First, by comparing both sets of estimates we may be able to learn something about whether achievement peer effects from new peers build up over time spent together, or fade as Year 5 achievement becomes a more distant proxy for current peer quality. (We know from the Year 7 estimates that there are almost immediate impacts, given that the Year 7 NAPLAN tests take place within the first half of the introductory secondary school year.) Second, NAPLAN tests in Year 9 take place closer to the age at which the equivalent tests in England, studied by LEA, take place. This, and the corresponding increase in ability tracking with grades, may make our Year 9 estimates more closely comparable with those of LEA. Of course we cannot explicitly separate the two potential sources of difference from one another, and there may also be other differences between Year 9 and Year 7 pupils that lead to heterogeneity in estimated peer effects. Also note that just two cohorts of Year 9 students are available for the analysis. Repeating the Year 7 analysis restricted to these two cohorts, however, suggests very similar patterns of peer effects as in the four-cohort case reported in Table 4.

The progression of the results in Table 8 from substantial to small, as controls for old peers, individual prior achievement and school*subject dummies are added, is similar to that for the Year 7 students in Table 4. In our preferred (sixth) specification, the magnitude of the statistically significant and positive average per effect is slightly smaller than that for Year 7 students, and we again find no evidence for a peer effect specifically from low-achieving peers. The positive impact from high-achieving peers which we reported in Table 4, however, is not there in Year 9; this effect actually becomes negative, although non-significant and sensitive to precise model specification. One interpretation could be that the average new peer effect neither substantially accumulates nor substantially declines over time; it is immediate or almost immediate, long-lasting, but not cumulative. In support of this, estimating (1) where Year 5 prior scores for individuals and peers are replaced by Year 7 prior scores yields zero peer effects for all three new peer variables. In other words, the impact of new peers is already reflected in the Year 7 own prior score. The disappearing impact of the top 10% of peers is consistent with a greater degree of ability tracking in Year 9, and brings our conclusions closer to those of LEA at least in this one respect.

As for the Year 7 results, these estimates are broadly robust to a number of different sample restrictions and other variations.[16] Estimates vary by gender (only girls are affected by their peers and only by the average new peer variable) and by school size (smaller and statistically

[16] Selected results are available from the authors on request.

insignificant for small schools). In both respects, but particularly the disappearance of the average peer effect for small schools, these results are closer to those of LEA. Smaller schools also suggest a statistically significant negative peer effect from low-achieving peers where the proportion of new peers is high, again more consistent with the overall conclusion of LEA. For large schools and for non-native English speakers the insignificant negative impact of high-achieving peers reported in Table 8 increases in magnitude and becomes statistically significant.

6. Conclusions

This paper provides evidence about the existence of achievement-related peer effects across multiple learning domains and school grades, but particularly in the first year of secondary schooling. It exploits measurements of peer quality prior to any social interaction and uses individual and school-subject fixed-effects for identification. This approach controls for common individual factors that influence achievement across domains, such as ability, motivation and background factors, along with school by subject effects common across individuals. It also avoids problems associated with "Angrist mechanics". Together, this provides a reasonable basis for interpreting the resulting peer effects as causal.

The estimates suggest that there are positive and significant average peer effects and that the proportion of top peers has an additional positive impact on individual achievement (the latter effect is there in the main estimates and in most specifications). These effects are small. Imagine two otherwise similar students – one with new peers who were 10 percentile rank points higher in their average peer score for a particular domain (just over a one standard deviation increase) than another student. This would translate into a 0.8 percentile rank gain for the student with better peers in Year 7, slightly smaller for a Year 9 student. This is equivalent to an improvement of about 0.05 of the within-student standard deviation of Year 7 test scores. The impact of the proportion of good peers an individual has is smaller again. Hattie (2009) views effect sizes in education of 0.2 of a standard deviation as small. Inasmuch as there is anything approaching consensus in the school achievement peer effects literature, it is probably that peer effects, if they exist at all, are small. Our results, which relate to peer achievement effects on individual achievement, are consistent with these effects being very small.

Our results are also consistent with existing evidence that peer effects may be heterogeneous along various dimensions. The average effects appear to be common for boys and girls, though the magnitude of the effect is marginally larger for females. The positive effect from the proportion of good students appears to be predominantly a female phenomenon. For both genders, the average new peer affect seems larger for those in the top half of the achievement distribution than those in the lower half, while the 'good' peers effect for females is stronger for lower performing students than for those in the top half of the prior achievement distribution.

In many respects our results seem to be the mirror image of those found in LEA, although they are closer to those of Gibbons and Telhaj (2015). We discuss a number of possible explanations for this, with the most likely being that the country, year group and subject mix studied by LEA mean their grade-level measures of peer achievement are poor measures of class-level peer achievement because of extensive within-grade ability tracking. The bottom line – perhaps no surprise – is that the institutions of the countries to which estimates relate appear to matter for the results. Conclusions regarding peer effects in school achievement are sensitive to context.

LEA conclude their paper with analysis of two "thought experiments" involving dimensions of classroom tracking, arguing their results are robust enough to support "insightful interventions targeting students' ability mix as a means to improve learning standards". Specifically, they come down broadly (but sensibly hesitantly) in favour of ability tracking in English secondary schools on efficiency grounds because this removes the negative effect of low-achieving peers on high achievers. Our results are at odds with their results, and we come down broadly (but similarly hesitantly) against further ability tracking in Australian schools on efficiency grounds to the extent that this reduces the potential for the positive impact within schools of high achievers on low achievers observed in the data. Having said that, it also seems likely that the peer-effects-related costs (or benefits) of extensive tracking in countries such as Australia are likely to be small.

We conclude with one final thought experiment relating to the roughly 12% of each cohort in the final year of public primary school who transit not into public secondary schools but private secondary schools, including elite private schools. The loss of these students induces concerns about the "residualisation" of low achievement and low socio-economic status students in public schools. Imagine that some intervention induced these students to stay predominantly in the public system and that any students who transferred were no longer disproportionately drawn from the top of the achievement distribution. In that case, both the proportion of new peers from the top and bottom deciles in public secondary schools would average 10%, while the average peer rank would be 50%. Would ability-related peer effects lead to a substantial achievement gain in the public school sector under this scenario? The mechanisms for such a gain are the positive impact of high achieving peers on others, particularly those in the bottom half of the ability distribution, and the average peer effect. This scenario would have only a small impact on student achievement in public schools, however, because the estimated peer effects are small. A 1.5 percentage point increase in the top group, multiplied by the 0.038 parameter from Table 4, and an increase in average peer quality of 2.5 percentage points combine to provide the equivalent of a 0.26 percentile improvement in average achievement across all public school students. This is less than 0.02 of the within-student standard deviation of Year 7 test scores and constitutes a tiny estimated effect, just one-tenth of the magnitude Hattie (2009) views as "small".

References

Altinok, N. and Kingdon, G. (2012). 'New evidence on class size effects: a pupil fixed effects approach.' Oxford Bulletin of Economics and Statistics Vol 74, 2, 207-234.

Angrist, J.D. (2014). 'The Perils of Peer Effects.' *Labour Economics*, 30, 98-108.

Angrist, J.D. and Lang, K. (2004). 'Does school integration generate peer effects? Evidence from Boston's Metco Program.' *American Economic Review*, 94 (5), 1613-1634.

Atkinson, A., Burgess, S., Gregg, P., Propper, C. and Proud, S. (2008). 'The impact of classroom peer groups on pupil GCSE results.' Working Paper 08/187, Centre for Market and Public Organisation, University of Bristol.

Bettinger, E.P. (2012). 'Paying to learn: the effect of financial incentives on elementary school test scores.' The Review of Economics and Statistics, vol. 94(3), 686-698.

Dee, T.S. (2005). 'A teacher like me: Does race, ethnicity, or gender matter?' American Economic Review 95 (2), 158–165.

Duflo, E., Dupas, P. and Kremer, M. (2011). 'Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya.' *American Economic Review*, 101 (5), 1739-1774.

Feld, J. and Zolitz, U. (2014). 'On the nature of peer effects in academic achievement.' Working Papers in Economics No. 596, University of Gothenburg.

Forgasz, H. (2010). 'Streaming for Mathematics in Years 7-10 in Victoria: An Issue of Equity?' *Mathematics Education Research Journal* 22 (1), 57-90.

Gibbons, S. and Telhaj, S. (2008). 'Peers and achievement in England's secondary schools.' SERC Discussion Paper No. 0001, London School of Economics.

Gibbons, S. and Telhaj, S. (2015). 'Peer effects: evidence from secondary school transition in England.' Oxford Bulletin of Economics and Statistics (forthcoming), DOI: 10.1111/obes.12095.

Goux, D. and Maurin, E. (2007). 'Close neighbours matter: neighbourhood effects on early performance at school.' *Economic Journal*, 117 (523), 1193-1215.

Hanushek, E.A., Kain, J.F., Markman, J.M. and Rivkin, S.G. (2003). 'Does peer ability affect student achievement?' *Journal of Applied Econometrics*, 18 (5), 527-544.

Hattie, J.A.C (2009), Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement, Routledge, Abingdon, UK.

Lavy, V. (2010). 'Do differences in school's instruction time explain international achievement gaps in math, science, and reading? Evidence from developed and developing countries.' NBER Working Paper 16227, National Bureau of Economic Research.

Lavy, V., Silva, O. and Weinhardt, F. (2012). 'The good, the bad, and the average: evidence on ability peer effects in schools.' *Journal of Labor Economics*, 30 (2), 367-414.

Lefgren, L. (2004). 'Educational peer effects and the Chicago public schools.' *Journal of Urban Economics*, 56, 169-191.

Manski, C. (1993). 'Identification of endogenous social effects: the reflection problem.' *Review of Economic Studies*, 60, 531-42.

Manski, C. (2000). 'Economic analysis and social interactions.' *Journal of Economic Perspectives*, 14(3), 115-136.

Mendolia, S., Paloyo, AR., Walker, I. (2016). 'Heterogeneous effects of high-school peers on educational outcomes.' IZA Discussion Paper 9795, Bonn: Institute for the Study of Labor.

Moffitt, R. (2001). ‘Policy interventions, low-level equilibria, and social interactions,’ in *Social Dynamics*, S. Durlauf and P. Young (eds.), Cambridge: MIT Press.

Sacerdote, B. (2011). ‘Peer effects in education: how might they work, how big are they and how much do we know thus far?’ In Hanushek, E., Machin, S. and Weossman. L. (eds.) *Handbook of the Economics of Education Volume 3*, Amsterdam: Elsevier.

Schwerdt, G. and Wuppermann, A.C. (2011). ‘Is traditional teaching really all that bad? A within-student between-subject approach.’ Economics of Education Review, 30, 2, 365-379.

Thomson, S., Lokan, J., Lamb, S. and Ainley, J., (2003) "Lessons from the Third International Mathematics and Science Study", Australian Council for Educational Research, http://research.acer.edu.au/timss_monographs/9

Vigdor, J. and Nechyba, T. (2007). ‘Peer effects in North Carolina public schools.’ in Schools and the equal opportunity problem, L. Woessmann and P. Peterson (eds.), Cambridge: MIT Press.

Figure 1: Sensitivity of estimated new peer effects to top and bottom peer group thresholds – Year 7 results

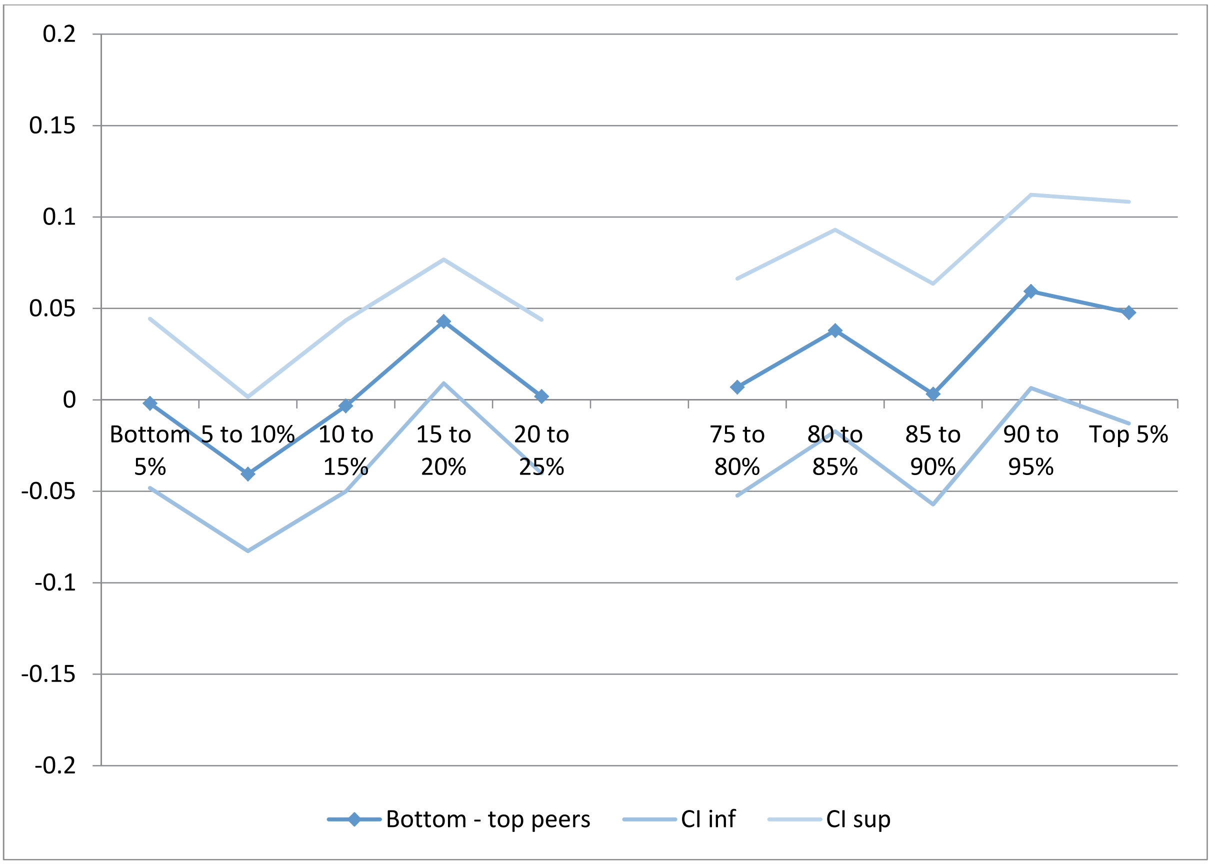

Table 1: Sample Means & Standard Deviations, Year 7 students in public schools

	All	"Regular students"	At least 1 subject in the top 10%	At least 1 subject in the bottom 10%
A. Pupil's outcomes				
Y5 percentile rank, grammar	48.653	47.249	75.083	21.616
	(28.509)	(24.145)	(19.666)	(19.835)
Y5 percentile rank, reading	48.390	46.509	75.374	21.797
	(28.472)	(23.919)	(19.745)	(20.046)
Y5 percentile rank, spelling	49.059	47.381	75.368	22.921
	(28.649)	(23.971)	(20.922)	(21.206)
Y5 percentile rank, writing	48.849	48.193	70.974	24.663
	(28.484)	(25.299)	(23.120)	(21.242)
Y7 percentile rank, grammar	50.605	49.798	81.094	20.307
	(28.576)	(21.707)	(16.955)	(18.686)
Y7 percentile rank, reading	50.724	49.982	80.844	20.310
	(28.582)	(21.578)	(17.462)	(19.345)
Y7 percentile rank, spelling	50.475	49.692	79.660	21.501
	(28.558)	(21.970)	(18.876)	(20.086)
Y7 percentile rank, writing	50.595	50.616	78.068	21.308
	(28.633)	(22.377)	(20.454)	(19.967)
B. Pupil's characteristics				
Male	0.518	0.491	0.469	0.638
	(0.500)	(0.500)	(0.499)	(0.481)
Speaks another language than English at home	0.228	0.206	0.290	0.220
	(0.420)	(0.404)	(0.454)	(0.414)
Indigenous	0.017	0.015	0.005	0.035
	(0.130)	(0.121)	(0.071)	(0.185)
Mother's occupation				
senior managmnt & qual. professionals	0.083	0.071	0.163	0.031
	(0.276)	(0.256)	(0.369)	(0.174)
Other business managers, arts/media/sportspersons and associate professionals	0.144	0.144	0.194	0.090
	(0.351)	(0.351)	(0.395)	(0.286)
Tradesmen/women, clerks and skilled office, sales and service staff	0.216	0.232	0.214	0.179
	(0.412)	(0.422)	(0.410)	(0.383)
Machine operators, hospitality staff, assistants, labourers and related workers	0.241	0.252	0.185	0.270
	(0.428)	(0.434)	(0.389)	(0.444)
Not in paid work in last 12 months	0.303	0.288	0.233	0.415
	(0.460)	(0.453)	(0.423)	(0.493)

(continued)	All	"Regular students"	At least 1 subject in the top 10%	At least 1 subject in the bottom 10%
Father's occupation				
senior managmnt & qual. professionals	0.081	0.068	0.166	0.026
	(0.273)	(0.252)	(0.372)	(0.160)
Other business managers, arts/media/sportspersons and associate professionals	0.151	0.151	0.213	0.086
	(0.358)	(0.358)	(0.410)	(0.280)
Tradesmen/women, clerks and skilled office, sales and service staff	0.208	0.223	0.200	0.177
	(0.406)	(0.417)	(0.400)	(0.382)
Machine operators, hospitality staff, assistants, labourers and related workers	0.217	0.223	0.162	0.258
	(0.412)	(0.416)	(0.368)	(0.438)
Not in paid work in last 12 months	0.113	0.105	0.092	0.155
	(0.317)	(0.307)	(0.290)	(0.362)
Mother's school education				
Year 9 or equivalent or below	0.094	0.087	0.045	0.162
	(0.292)	(0.282)	(0.208)	(0.368)
Year 10 or equivalent	0.186	0.193	0.107	0.254
	(0.389)	(0.394)	(0.309)	(0.435)
Year 11 or equivalent	0.218	0.238	0.162	0.228
	(0.413)	(0.426)	(0.368)	(0.419)
Year 12 or equivalent	0.477	0.459	0.666	0.328
	(0.499)	(0.498)	(0.472)	(0.469)
Father's school education				
Year 9 or equivalent or below	0.087	0.084	0.048	0.137
	(0.282)	(0.278)	(0.214)	(0.343)
Year 10 or equivalent	0.169	0.181	0.113	0.198
	(0.375)	(0.385)	(0.317)	(0.399)
Year 11 or equivalent	0.164	0.178	0.142	0.153
	(0.370)	(0.382)	(0.349)	(0.360)
Year 12 or equivalent	0.338	0.317	0.522	0.200
	(0.473)	(0.465)	(0.500)	(0.400)
Mother's non school education				
No non school qualification	0.371	0.377	0.254	0.478
	(0.483)	(0.485)	(0.435)	(0.500)
Certificate I to IV (incl. trade certificate)	0.263	0.280	0.207	0.276
	(0.440)	(0.449)	(0.405)	(0.447)
Advanced diploma/Diploma	0.128	0.132	0.151	0.093
	(0.334)	(0.338)	(0.358)	(0.291)
Bachelor degree or above	0.166	0.137	0.335	0.063
	(0.372)	(0.344)	(0.472)	(0.243)

(continued)	All	"Regular students"	At least 1 subject in the top 10%	At least 1 subject in the bottom 10%
Father's non school education				
No non school qualification	0.250	0.257	0.184	0.303
	(0.433)	(0.437)	(0.387)	(0.460)
Certificate I to IV (incl. trade certificate)	0.262	0.282	0.224	0.252
	(0.440)	(0.450)	(0.417)	(0.434)
Advanced diploma/Diploma	0.083	0.084	0.111	0.052
	(0.276)	(0.277)	(0.315)	(0.222)
Bachelor degree or above	0.127	0.100	0.280	0.039
	(0.333)	(0.300)	(0.449)	(0.194)
Size of peer group (in Y7)	176	175	192	160
	(78)	(78)	(78)	(75)
N	107,245	59,534	24,333	23,378

Table 2: Average Peer rankings

	Mean	Std	Min	Max
Percentage of new peers in Y7	86.32	17.67	3.33	99.75
A. Average of new peers' Y5 percentile rank				
in grammar	47.72	9.60	1.07	98.41
in reading	47.51	9.79	1.75	93.41
in spelling	48.17	9.30	1.45	92.10
in writing	47.77	9.45	1.05	88.84
B. Percentage of new peers' Y5 percentile rank in top 10%				
in grammar	8.64	6.03	0	100
in reading	8.47	6.31	0	100
in spelling	9.01	6.43	0	74.06
in writing	8.54	6.26	0	50
C. Percentage of new peers' Y5 percentile rank in bottom 10%				
in grammar	11.27	7.69	0	100
in reading	11.33	7.66	0	100
in spelling	11.15	7.22	0	100
in writing	11.17	7.46	0	100

Table 3: Analysis of variance of peer rankings

	Mean	Overall Std	Between Std	Within Std
Y7 percentile rank	50.60	28.59	24.50	14.92
Y5 percentile rank	48.74	28.53	24.08	15.51
Average of new peers' Y5 percentile rank	47.79	9.54	8.98	3.18
Percentage of new peers' Y5 percentile rank in top 10%	8.66	6.26	5.59	2.79
Percentage of new peers' Y5 percentile rank in bottom 10%	11.23	7.51	6.70	3.42

Table 4: Regression results – New peers and Year 7 achievement

		OLS			Student fixed effect		
		Average	Bottom 10%	Top 10%	Average	Bottom 10%	Top 10%
1. Peer variables entered separately, no controls for old peers nor own prior score							
	Coefficient	0.470***	-0.425***	0.627***	0.363***	-0.168***	0.280***
	se	(0.015)	(0.030)	(0.024)	(0.022)	(0.019)	(0.023)
	N	461,861	461,861	461,861	461,861	461,861	461,861
2. Peer variables entered together, no controls for old peers nor own prior score							
	Coefficient	0.404***	0.016	0.140***	0.315***	-0.018	0.080***
	se	(0.024)	(0.024)	(0.022)	(0.026)	(0.016)	(0.021)
	N	461,861	461,861	461,861	461,861	461,861	461,861
3. Peer variables entered together, controls for old peers but not for own prior score							
	Coefficient	0.288***	0.006	0.085***	0.267***	-0.019	0.067***
	se	(0.022)	(0.021)	(0.021)	(0.025)	(0.015)	(0.021)
	N	418,821	418,821	418,821	418,821	418,821	418,821
4. Peer variables entered together, controls for old peers and for same-subject own prior score interacted with subject dummies							
	Coefficient	0.153***	-0.001	0.055***	0.179***	-0.014	0.042**
	se	(0.018)	(0.016)	(0.018)	(0.021)	(0.014)	(0.019)
	N	397,896	397,896	397,896	397,896	397,896	397,896
5. Peer variables entered together, controls for old peers and for cross-subject own prior scores interacted with subject dummies							
	Coefficient	0.135***	0.000	0.037**	0.120***	-0.013	0.034*
	se	(0.017)	(0.015)	(0.017)	(0.019)	(0.014)	(0.018)
	N	393,677	393,677	393,677	393,677	393,677	393,677
6. Peer variables entered together, controls for old peers and for cross-subject own prior scores interacted with subject dummies + school*subject dummies							
	Coefficient	0.133***	0.000	0.044**	0.081***	-0.005	0.038**
	se	(0.019)	(0.016)	(0.019)	(0.021)	(0.015)	(0.019)
	N	393,677	393,677	393,677	393,677	393,677	393,677

Controls: GRM RDG SPL GRMXsex RDGXsex SPLXsex Also, the OLS models include individual controls for: sex, language spoken at home, indigenous status, mother's and father's occupation, mother's and father's school & non-school education. Robust standard errors clustered at the school level. ***/**/* denotes statistical significance at the 99%/95%/90% levels.

Table 5: Placebo Tests, Year 7, (i) replacing own score with own Year 5 score, with subject/school dummies, and (ii) replacing subject peer scores with those from random subjects, individual fixed effects model

		Average	Bottom 10%	Top 10%
1. Placebo test: Yr5 outcome, peer variables entered together, controls for old peers but not for own prior score + school*subject dummies				
	Coefficient	-0.009	-0.026*	0.030
	se	(0.024)	(0.013)	(0.021)
	N	397,896	397,896	397,896
2. Placebo test: Random peer values entered for subjects, peer variables entered together, controls for old peers and for cross-subject own prior scores interacted with subject dummies + school*subject dummies				
	Coefficient	-0.010	0.001	0.000
	se	(0.008)	(0.006)	(0.008)
	N	397,896	397,896	397,896

Controls: GRM RDG SPL GRMXsex RDGXsex SPLXsex Robust standard errors clustered at the school level. ***/**/* denotes statistical significance at the 99%/95%/90% levels.

Table 6: Regression results – Year 7 new peer effects by gender and ability

	Boys			Girls		
	Average	Bottom 10%	Top 10%	Average	Bottom 10%	Top 10%
A. Overall effect						
	0.074***	-0.007	0.028	0.087***	-0.006	0.051**
	(0.024)	(0.019)	(0.024)	(0.027)	(0.018)	(0.022)
B. By ability (wrt YR5 percentile rank)						
Percentile rank < 20	0.037	-0.021	0.057*	0.065**	-0.047**	0.061*
	(0.027)	(0.020)	(0.034)	(0.031)	(0.022)	(0.036)
Percentile rank 20-35	0.066**	-0.021	0.055	0.075**	-0.019	0.080**
	(0.030)	(0.024)	(0.039)	(0.034)	(0.026)	(0.039)
Percentile rank 35-50	0.014	-0.046	0.077*	0.072**	-0.003	0.076**
	(0.032)	(0.028)	(0.042)	(0.036)	(0.028)	(0.038)
Percentile rank 50-65	0.131***	0.020	-0.026	0.100***	0.026	0.075**
	(0.035)	(0.032)	(0.041)	(0.036)	(0.029)	(0.036)
Percentile rank 65-80	0.111***	-0.003	-0.009	0.108***	0.010	0.016
	(0.037)	(0.028)	(0.037)	(0.033)	(0.027)	(0.034)
Percentile rank >= 80	0.124***	0.047*	-0.018	0.095***	0.001	0.023
	(0.033)	(0.028)	(0.033)	(0.030)	(0.022)	(0.030)
N	201,566	201,566	201,566	192,111	192,111	192,111

Controls: GRM RDG SPL GRMXsex RDGXsex SPLXsex. Student FE model with school subject dummies and all other controls. Robust standard errors clustered at the school level. ***/**/* denotes statistical significance at the 99%/95%/90% levels.

Table 7: Regression results – Year 7 new peer effects by school size and proportion of new peers

	Small schools			Large schools		
	Average	Bottom 10%	Top 10%	Average	Bottom 10%	Top 10%
A. Overall effect						
	0.039*	-0.008	0.036*	0.282***	-0.019	0.018
	(0.024)	(0.015)	(0.021)	(0.069)	(0.075)	(0.059)
B. By quintiles of the proportion of new peers						
Proportion of new peers < Q1	0.006	0.003	0.036	0.285***	-0.045	0.021
	(0.027)	(0.016)	(0.026)	(0.070)	(0.078)	(0.074)
Proportion of new peers >=Q1 & <Q2	0.065*	-0.025	0.052	0.245***	-0.010	0.050
	(0.038)	(0.031)	(0.049)	(0.081)	(0.099)	(0.070)
Proportion of new peers >=Q2 & <Q3	0.087**	-0.023	0.002	0.360***	-0.009	0.001
	(0.036)	(0.033)	(0.047)	(0.088)	(0.097)	(0.078)
Proportion of new peers >=Q3 & <Q4	0.091**	-0.054	0.021	0.332***	0.020	-0.021
	(0.039)	(0.038)	(0.047)	(0.091)	(0.117)	(0.079)
Proportion of new peers >=Q4	0.052	-0.101**	0.107*	0.299***	-0.005	0.033
	(0.046)	(0.048)	(0.065)	(0.098)	(0.107)	(0.084)
N	193,395	193,395	193,395	200,282	200,282	200,282

Controls: GRM RDG SPL GRMXsex RDGXsex SPLXsex. Student FE model with school subject dummies and all other controls. Robust standard errors clustered at the school level. ***/**/* denotes statistical significance at the 99%/95%/90% levels.

Table 8: Regression results – New peers and Year 9 achievement

	OLS			Student fixed effect		
	Average	Bottom 10%	Top 10%	Average	Bottom 10%	Top 10%
1. Peer variables entered separately, no controls for old peers nor own prior score						
Coefficient	0.429***	-0.365***	0.577***	0.336***	-0.194***	0.240***
se	(0.024)	(0.044)	(0.036)	(0.026)	(0.025)	(0.033)
N	204,320	204,320	204,320	204,320	204,320	204,320
2. Peer variables entered together, no controls for old peers nor own prior score						
Coefficient	0.387***	0.053	0.141***	0.278***	-0.050**	0.068**
se	(0.041)	(0.038)	(0.037)	(0.032)	(0.024)	(0.029)
N	204,320	204,320	204,320	204,320	204,320	204,320
3. Peer variables entered together, controls for old peers but not for own prior score						
Coefficient	0.286***	0.046	0.108***	0.233***	-0.051**	0.061**
se	(0.037)	(0.032)	(0.035)	(0.031)	(0.024)	(0.029)
N	186,326	186,326	186,326	186,326	186,326	186,326
4. Peer variables entered together, controls for old peers and for same-subject own prior score interacted with subject dummies						
Coefficient	0.152***	0.036	0.101***	0.136***	-0.033	0.043
se	(0.030)	(0.025)	(0.031)	(0.028)	(0.022)	(0.029)
N	178,123	178,123	178,123	178,123	178,123	178,123
5. Peer variables entered together, controls for old peers and for cross-subject own prior scores interacted with subject dummies						
Coefficient	0.141***	0.034	0.080***	0.078***	-0.028	0.030
se	(0.029)	(0.024)	(0.029)	(0.026)	(0.021)	(0.027)
N	176,371	176,371	176,371	176,371	176,371	176,371
6. Peer variables entered together, controls for old peers and for cross-subject own prior scores interacted with subject dummies + school*subject dummies						
Coefficient	0.135***	0.066***	-0.023	0.068**	0.015	-0.043
se	(0.034)	(0.023)	(0.032)	(0.031)	(0.021)	(0.031)
N	176,371	176,371	176,371	176,371	176,371	176,371

Controls: GRM RDG SPL GRMXsex RDGXsex SPLXsex. Student FE model with school subject dummies and all other controls. Robust standard errors clustered at the school level. ***/**/* denotes statistical significance at the 99%/95%/90% levels.