Received: 7 August 2017 | Revised: 26 September 2018

DOI: 10.1002/jae.2674

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Sibling spillover effects in school achievement

Cheti Nicoletti[1,2] | Birgitta Rabe[2]

[1]Department of Economics and Related Studies, University of York, York, UK

[2]Institute for Social and Economic Research, University of Essex, Colchester, UK

Correspondence
Birgitta Rabe, Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ, UK.
Email: brabe@essex.ac.uk

Summary

This paper provides empirical evidence on sibling spillover effects in school achievement using administrative data on 230,000 siblings in England. We extend previous strategies to identify peer effects by exploiting the variation in school test scores across subjects observed at ages 11 and 16 as well as variation in peer quality between siblings. We find a statistically significant positive spillover effect from the older to the younger sibling. Sibling spillovers account for a non-negligible proportion of the attainment gap between low- and higher income pupils in England.

1 | INTRODUCTION

In this paper we study the extent to which school achievements of an older sibling affect the school outcomes of their younger sibling. Siblings usually spend a lot of time together, and it is likely that they influence each other's school outcomes through teaching and help with homework, modeling academic behaviors, educational aspirations and values, and sharing information. If there is a large role of siblings for educational outcomes, this means that investments into children by parents or schools may be amplified through sibling spillovers, suggesting that there are externalities of parental and public investments into children through their positive effects on siblings. Moreover, sibling spillovers may have implications for inequality in outcomes of children from different socioeconomic backgrounds—for example, if high-income children are more likely to benefit from the transmission of good behaviors than low-income children are.

While the economic literature recognizes the important role of parent–child interactions for child development, the role of sibling interactions has received much less attention. There is an extensive literature providing evidence on the existence of school peer effects in cognitive ability (for reviews, see; Epple & Romano, 2011; Sacerdote, 2011). However, interactions between siblings are arguably the most frequent and relevant interactions a child may have with other children, and they could therefore result in substantially larger spillover effects. Sibling correlations in socioeconomic and educational outcomes have been used to describe the importance of family background (see; Björklund & Salvanes, 2011; Lindahl, 2011; Mazumder, 2008; Nicoletti & Rabe, 2013; Raaum, Salvanes, & Sørensen, 2006), but they cannot quantify the spillover effect attributable to sibling interactions. It is only recently that economists have begun to look at the causal effect of interactions between siblings on educational outcomes. For example, evidence on causal sibling spillover effects on high school graduation by age 19, years of schooling, and subject and school choices is provided by Oettinger (2000), Qureshi (2018a), Joensen and Nielsen (2017), and Dustan (2018).

We add to this literature by providing, to our knowledge for the first time, empirical evidence on the extent to which the school achievement of a child is transmitted to his or her younger sibling. More precisely, we estimate the sibling spillover effect of a child's school test scores at age 16 on her younger sibling's test scores at the same age. By focusing on spillover effects in compulsory study subjects (English and Math), we are able to capture sibling influence on skills, effort, and motivation rather than on subject choice. Apart from adding to the emerging literature on sibling spillovers in education,

we make a methodological contribution by proposing a new strategy to identify sibling spillover effects that does not rely on policy reforms and therefore can be applied in other contexts. Further, we investigate the impact of sibling spillovers for children by family background, showing that a non-negligible proportion of the attainment gap between children from low- and higher income families can be explained by spillover effects between siblings.

To assess the role of interactions between siblings in the transmission of cognitive abilities we aim at producing an estimate of the sibling spillover effect that is cleaned as far as possible of indirect effects caused by parental behavioral responses or other confounding factors.[1] We use administrative data on the whole population of children in state schools in England, which allows us to identify siblings and school peers and to observe for each child outcomes of externally marked national tests in mathematics and English at several points in time. We use the variation of school test scores across subjects to eliminate individual fixed effects. Furthermore, to check whether there is any endogeneity caused by unobserved subject-specific inputs, we instrument the older sibling's school test score with the predetermined school performance of the older sibling's peers.

Simply regressing a child test score on the older sibling's corresponding test score would not produce a consistent estimate of the sibling spillover effect, because the estimated sibling association would be in part explained by similarities in inherited abilities, in school and family investments and characteristics, and in the shared environment siblings are exposed to. To clean the sibling association in test scores of these correlated observed and unobserved factors, we make use of test scores at the end of compulsory schooling (at about age 16) in mathematics and English. We regress a child's test score on her older sibling's test score using within-pupil between-subject estimation—that is, estimating child fixed effects.[2] The two main gains of this fixed-effect estimation are that it allows us to (i) control for the younger child's unobservable average ability and other characteristics that are invariant across the subjects and may confound the spillover effect because they are similar between siblings, and (ii) clean the sibling spillover effect of the impact of investments by schools and parents between siblings that do not vary across subjects. Further, to account for subject-specific school characteristics we rely on school-by-cohort-by-subject fixed effects for the younger sibling.

To further take account of *subject-specific* skills that are acquired from parents through family investments and/or inheritance and shared by siblings, we implement a two-stage least square (2SLS) estimation by instrumenting the older sibling's test scores at age 16 using the average test scores at age 11 of her schoolmates. Our instrumental variable strategy exploits idiosyncratic differences in peer group quality across schools and/or across cohorts. We will present sensitivity analyses to show that endogenous school choice is not driving the variation in the older sibling's school peers' achievements. To avoid reverse causality running from the older sibling to her schoolmates in the first-stage regression of the 2SLS estimation, we follow Gibbons and Telhaj (2016) and consider only the performance of new peers that the older sibling first encountered in secondary school and use the new peers' prior test scores obtained in primary school to measure attainment (see also Lavy et al., 2012). We run a number of specification and falsification tests to assess the relevance and validity of our instrument.

Our peer identification strategy is similar to that adopted by Kelejian and Prucha (1998), Lee (2003), Bramoullé, Djebbari, and Fortin (2009), De Giorgi, Pelizzari, and Redaelli (2010), De Giorgi, Fredriksen, and Pistaferri (2016), and Nicoletti, Salvanes, and Tominey (2018), and is based on the presence of some intransitivity in the network of peers. Intransitivity occurs if a person interacts with her peers but not with all of the peers of her peers. In our application we have intransitivity because we assume that the older sibling's schoolmates do not interact directly with the younger sibling. This implies that, while the older sibling's test scores can be affected directly by her schoolmates' results, there is no effect from the older sibling's schoolmates on the younger sibling (other than indirectly through the older sibling). We scrutinize this identifying assumption by performing sensitivity checks on the data, for example by excluding from the estimation sample schoolmates of the older sibling who live in the same area and might therefore interact directly with the younger sibling and by implementing falsification tests.

Our estimation strategy takes account of the three well-known identification issues in peer effect estimation (Brock & Durlauf, 2001; Dolton, 2017; Manski, 1993, 2000; Moffitt, 2001), which are issues of correlated omitted variables, reflection (simultaneity), and endogenous peer membership. After controlling for these issues we identify a spillover effect between

[1]There are some recent papers that have estimated the total policy effect on a child's educational outcomes of conditions or policy reforms affecting his or her siblings (see; Breining, 2014; Breining, Daysal, Simonsen, & Trandafir, 2015; Fletcher, Hair, & Wolfe, 2012; Qureshi, 2018b). The estimation of such total spillover effects, which include both the causal effect and the indirect effect, through behavioral responses by parents for example, is of policy interest, but it is not possible to generalize the indirect effect to other contexts.

[2]This estimation is similar in spirit to the within-pupil between-subject estimation used by Dee (2005, 2007), Clotfelter, Ladd, and Vigdor (2010), and Slater, Davies, and Burgess (2010), and it has been used by Lavy, Silva, and Weinhardt (2012) to estimate school peer effect on test scores.

siblings, which is what Manski (1993) defines as an endogenous peer effect—that is, a spillover effect from the older to the younger sibling net of any correlated omitted variables or net of what Manski calls contextual and correlated effects.

We control for correlated omitted variables that are invariant across subjects through the inclusion of the child fixed effect, whereas we control for correlated omitted variables, which are school, cohort, and subject specific thanks to the school-by-cohort-by-subject fixed effect. We cannot exclude the existence of a reflection issue—that is, the existence of a spillover effect going from the younger to the older sibling[3]; but this reverse causality is unlikely to occur in our application because the younger sibling's age 16 test score is in the future with respect to the corresponding older sibling's test score at age 16. The instrumental variable estimation allows us to check if there is any endogeneity bias caused by this reflection issue or by omitted family inputs and characteristics that vary across subjects. We find that there is no residual endogeneity left. Endogenous peer membership may occur if the likelihood to interact with peers depends on unobserved characteristics that also affect the test score. Peers are defined as children belonging to the same family (siblings) or school–cohort (school peers), so the likelihood to form interactions depends on the selection into the family and into the school. There may be selection into schools based on unobserved genetic traits and covariates. To control for the potential correlation between the subject-specific test score of a child's school peers and of his or her older sibling's schoolmates, we include the school-by-cohort-by-subject fixed effect of the younger sibling. In addition, we control for the child fixed effect and therefore for any unobserved genetic traits and covariates that do not vary across subjects.

We find that an increase of one standard deviation (SD) in a child's test score at age 16 leads to a statistically significant increase in the corresponding test score observed for his or her younger sibling of about 11% of a SD. This means that for each exam grade improvement of the older sibling—for example, from a B to an A—the younger sibling's exam marks increase by about 10% of a grade, which is equivalent to the impact of increasing yearly per pupil school expenditure in the younger sibling's school by around £1,000 (see; Nicoletti & Rabe, 2018). Clearly, there are externalities from investing in children's learning that have so far been mostly overlooked.

Heterogeneity analysis reveals that spillover effects are strongest for children whose older siblings are top achievers, and lowest for children whose older siblings perform badly. Interestingly, this pattern holds across families from different socioeconomic backgrounds. However, as disadvantaged children have a higher likelihood of having a badly performing older sibling than do affluent children, they benefit from the positive transmission less often. In fact, we find that children in low-income families (proxied by their eligibility for free school meals), on average, experience negative effects on their attainment as a result of sibling spillovers, whereas children in higher income families benefit on average. We estimate that about 8.4% of the attainment gap between children entitled and not entitled to free lunches (which is 61% of a SD) can be attributed to sibling spillover effects.

The remainder of this paper unfolds as follows. The next section lays out our identification strategy, and Section 3 introduces our data set. Section 4 presents our empirical results, including robustness checks and the estimation of heterogeneous spillover effects by subgroups. Section 5 concludes.

2 | IDENTIFICATION STRATEGY

To identify the sibling spillover effect on test scores at the end of compulsory schooling (at about age 16) we consider the following value-added model[4]:

$$Y_{1,isqt,16} = \alpha + Y_{1,isqt,11}\rho + Y_{2,is'qt',16}\gamma + \mathbf{I}^{F}_{1,i}\boldsymbol{\beta}_{1,F} + \mathbf{I}^{S}_{1,ist}\boldsymbol{\beta}_{1,S} + \mathbf{X}_{1,i}\boldsymbol{\beta}_{1,X} + \mathbf{Z}_{1,isqt}\boldsymbol{\beta}_{1,Z} + \mu_{sqt} + \mu_{1,i} + e_{1,isqt,16}, \quad (1)$$

where

- $Y_{1,isqt,16}$ is the age 16 test score of the younger child of the sibling-pair i, in school s and subject q, who belongs to the cohort t;[5]
- $Y_{1,isqt,11}$ is the corresponding test score at age 11;
- $Y_{2,is'qt',16}$ is the test score at age 16 of the older sibling, who might have attended a different school s' and belongs to a different cohort t';
- $\mathbf{I}^{F}_{1,i}$ is a vector of family inputs in the younger child of the sibling-pair i between age 11 and 16, which are not subject specific;

[3]Ewin Smith (1993) suggests that the cognitive abilities of older children might improve thanks to teaching younger siblings.
[4]See Todd and Wolpin (2003) for a definition.
[5]Two students belong to the same school cohort if they began school in the same year. We do not consider twins or siblings whose age gap is such that they begin school in the same year.

- $\mathbf{I}^{S}_{1,ist}$ is a vector of school inputs;
- $\mathbf{X}_{1,i}$ is a row vector of other child, household, and school characteristics, which are not direct inputs in the development of a child's subject specific skills but may affect them;
- $\mathbf{Z}_{1,isqt}$ is a row vector of other characteristics that can change across individual, subject, cohort, and school, and includes the younger sibling's subject-specific school achievements averaged over her school-by-cohort peers, as defined in more detail later' and the interaction between gender and subject;
- μ_{sqt} is an unobserved heterogeneity component capturing all other omitted school inputs that vary by school, subject, and cohort, but not across children in the same school–cohort;
- $\mu_{1,i}$ is an individual unobserved component capturing all other younger child's unobservables that do not vary across subjects and school–cohorts;
- $e_{1,isqt,16}$ is an idiosyncratic error term that is assumed to be identically and independently distributed with mean zero and homoskedastic.

In this model ρ measures the persistence in test scores between age 11 and 16; γ is our main parameter of interest, which measures the spillover effect from the older to the younger sibling; $\boldsymbol{\beta}_{1,F}$ and $\boldsymbol{\beta}_{1,S}$ are column vectors with the productivities of family and school investments; $\boldsymbol{\beta}_{1,X}$ and $\boldsymbol{\beta}_{1,Z}$ are column vectors with the effects of the remaining explanatory variables $\mathbf{X}_{1,i}$ and $\mathbf{Z}_{1,isqt}$, and α is the intercept. We observe for each sibling-pair their test scores in mathematics and English, so that q takes value 1 for mathematics and 2 for English.

Identifying the causal spillover effect in test scores from the older to the younger sibling, γ, is challenging because of three main issues: (i) unobserved correlated effects—that is, unobserved common characteristics of two siblings that may explain their similar test scores; (ii) the reflection problem—that is, reverse causality; (iii) the endogeneity of the network—that is, nonrandom sorting of individuals into groups.

To address (i), we control for unobserved child-specific endowments and characteristics that do not vary across subjects but that could be similar between siblings by transforming Model 1 in deviations from the mean across subjects—that is, we transform the dependent variable in $\text{Dev}Y_{1,isqt,16} = Y_{1,isqt,16} - \sum_{j=1}^{2} Y_{1,isjt,16}/2$ and we apply an analogous transformation to all right-hand-side variables, leading to

$$\text{Dev}Y_{1,isqt,16} = \text{Dev}Y_{1,isqt,11}\rho + \text{Dev}Y_{2is'qt',16}\gamma + \text{Dev}\mathbf{Z}_{1,isqt}\boldsymbol{\beta}_{1,Z} + \text{Dev}\mu_{sqt} + \text{Dev}e_{1,isqt,16}. \quad (2)$$

This transformation is equivalent to controlling for child fixed effects and eliminates from the equation all inputs that do not vary across subjects, as well as the unobserved child endowment, $\mu_{1,i}$. This comprises cognitive and noncognitive abilities and health, which could be similar between siblings, therefore confounding the sibling spillover effect.[6]

However, the child fixed effect is unable to capture unobserved characteristics that do vary by subject. In particular, we might be concerned about subject-specific *school* investments that are shared by siblings because they attend schools with similar (or indeed the same) characteristics that are unobserved by us. We partial out shared subject-specific school background by using school-by-cohort-by-subject fixed effects (for the younger sibling) that control for μ_{sqt}—that is, for unobserved subject-specific school investments and characteristics for the cohort t.

The issue of unobserved subject-specific *family* investments and inherited skills is more challenging. By controlling for the lagged test score—that is, the test score in subject q at age 11, we estimate a spillover effect that is largely purged of the influence of such family characteristics up to the age of 11. To control also for the effect of any unobserved subject-specific characteristics between ages 11 and 16, we adopt instrumental variable estimation where we instrument the subject-specific test scores of the older sibling at age 16 using the average attainment of the school-by-cohort peers of the older sibling. The endogeneity test of the IV estimates will allow us to test whether, after controlling for child fixed effects and school-by-cohort-by-subject fixed effects, there is residual endogeneity of the older sibling's test score.

We need to make sure that our instrumental variable estimation is not affected by reverse causality (the reflection problem, issue (ii)). Therefore, we adopt the strategy used by Gibbons and Telhaj (2016) and Lavy et al. (2012), who measure peers' ability using prior achievements in end-of-primary-school national tests at age 11 but only considering *new* peers that a pupil (in our case the older sibling) encounters for the first time in secondary school. In the compulsory transition from primary to secondary school a major reshuffling of pupils takes place, so that on average students meet 86% new peers. This instrument is immune to reflection problems because the older sibling's test score at age 16 cannot

[6]If we chose to include controls pertaining to the older sibling in the model ($I^{F}_{2,it}, I^{S}_{2,ist}, X_{2,i}$), these would also be eliminated with this transformation. Note that instead of deviations from the mean we could take first differences between mathematics and English, which would yield identical estimates.

affect her new school peers' test scores at age 11 and because the younger sibling's test score does not affect the older sibling's new school peers' test scores at age 11.[7]

We instrument the subject-specific test scores of the older sibling at age 16 using the average of $\text{Dev}Y_{js'qt',11}$ (measured in attainment percentiles) over the *new* school-by-cohort peers of the older sibling, which we call $\text{New}M\text{Dev}Y_{2,s'qt',11}$. In Equation (2), apart from taking account of both child fixed effects and school-by-cohort-by-subject fixed effects of the younger sibling, we also control for the $\text{Dev}\mathbf{Z}_{1,isqt}$, which includes the younger sibling's subject-specific attainment percentiles averaged over the new school-by-cohort peers of the younger sibling; that is, $\text{New}M\text{Dev}Y_{1,sqt,11}$. In this way, the instrument captures whether the older sibling's new school–cohort mates were relatively better in a specific subject than the younger sibling's new school–cohort mates, after partialling out the effect of any older sibling's mates who have a younger sibling in the same school–cohort as our reference sibling pair. The variation in the instrument comes from idiosyncratic differences in average subject-specific peer quality between groups of school peers within and across schools and cohorts. These differences can occur, for example, because of changes in the quality of teaching in a specific subject (e.g., because of teacher turnover) or in the composition of the new school–cohort mates in terms of inherent subject-specific abilities.[8]

To be a valid instrumental variable $\text{New}M\text{Dev}Y_{2,s'qt',11}$ must be uncorrelated with any unobserved school or student characteristics that affect the younger sibling's test results. In particular, we are concerned that siblings from one family might sort into secondary schools according to subject preferences that are shared by peers—for example, similar quality of teachers in a particular subject or peers with similar subject-specific abilities. This is a concern about the endogeneity of the network (issue (iii)). In England, students generally attend primary and secondary schools according to residence-based school catchment areas. However, parents are free to apply for school places outside their catchment area, subject to availability of a school place. While this is less common at primary school stage, at secondary school stage students diversify more according to school preference.

Our estimation approach takes care of the issue of sorting in several ways. By including school–by–cohort–by–subject fixed effects for the younger sibling we control for persistent subject-specific school unobservables. The subject-specific average test results of peers of the younger sibling, $M\text{Dev}Y_{1,sqt,11}$, which are likely to be correlated with the instrument in the presence of sorting, also gets wiped out by this fixed effect. Moreover, we explicitly control for the prior average test results of the younger sibling's new school peers, $\text{New}M\text{Dev}Y_{1,sqt,11}$ which can also be correlated with the instrument. Finally, there could be variation in the instrumental variable caused by the fact that older siblings in the same secondary school attended different primary schools due to endogenous sorting. Because secondary schools are much larger than primary schools in England, several primary schools usually feed into one secondary school, resulting in the majority of peers (86% on average) to be new to each pupil. This means that there will be a large overlap in new peers even between two students coming from different primary schools, mitigating the concern that endogenous school choice drives variation in our instrument. We discuss robustness checks addressing potentially endogenous school choice in Section 4.2.

Another identifying assumption is that a student can be affected by the test scores of the new school peers of her sibling only through her sibling. This assumption could be invalid if there is direct interaction between the older sibling's new school mates and the younger sibling, for example. We discuss this and other possible threats to identification in Section 4.2 and present a number of robustness checks. For example, we exclude the older sibling's school peers who live in the same neighborhood from the computation of $\text{New}M\text{Dev}Y_{2,s'qt',11}$ to assess whether possible interaction within a neighborhood might affect results. We conclude from these checks that our estimated sibling spillover effect holds across a number of specifications.

Finally, in order to isolate the causal effect of the older sibling's attainment on the younger, we also need to assume that there are no behavioral responses by parents. Our identification strategy is more immune to this problem than previous papers that rely on policy reforms that affect one sibling and not the other—for example, an increase in the school leaving age. In the case of policy reforms we expect parents to adjust the allocation of their investments between siblings so that the instrumental variable is not independent of these investments. Of course, parents could also react to our instrument, which captures whether the older sibling's new school–cohort mates were relatively better in a specific subject than the younger sibling's school–cohort mates, after partialling out the effect of older sibling mates who have a younger sibling in the same school–cohort as our reference sibling pair. Parents are unlikely to perceive this type of variation, especially

[7]The variation in the new peers' test scores for older siblings who attend the same secondary school but come from different primary schools is arguably difficult to interpret, but such type of variation does not seem to be driving our results. Indeed, instrumental variable estimates based on all peers rather than new peers remain similar.

[8]Partialling out the latter effect avoids any influence of the older sibling's mates through their younger siblings who are in the same school–cohort as our reference child.

if it is across cohorts within schools. It might be easier to observe if the variation in the instrument is across schools, but we have checked that our results hold for a sample of siblings attending the same school. Because the spillover effect on the younger sibling is measured several years after the older sibling is first exposed to her new peers, it may be possible, however, that parents observe and respond to any contemporaneous effect on the younger sibling over the years.

3 | DATA

The empirical analysis is based on the National Pupil Database (NPD), which is available from the English Department for Education and has been widely used for education research. The NPD is a longitudinal register data set for all children in state schools in England, covering roughly 93% of English students. It combines student-level attainment data with student characteristics as they progress through primary and secondary school.

3.1 | Educational system in England

Full-time education is compulsory for all children aged between 5 and 16, with most children attending primary school from age 5 to 11 and secondary school from age 11 to 16. There is no grade retention in the UK and virtually all children attend school at the expected age. The education during these years is divided into four Key Stages. Students take externally marked National Curriculum tests at the end of Key Stages 2 and 4. Until recently, such national tests were also carried out at Key Stages 1 and 3 but today progress at these stages is examined via individual teacher assessment.

Key Stage 2 National Curriculum tests are taken at the end of primary school, usually at age 11. Pupils take tests in the three core subjects of English, mathematics, and science. Key Stage 4 tests are taken at age 16 at the end of compulsory schooling. Pupils enter General Certificate of Secondary Education (GCSE) or equivalent vocational or occupational exams at this stage. They decide which GCSE courses to take and, because English, mathematics, and science are compulsory study subjects, virtually all students take GCSE examinations in these topics, plus others of their choice, with a total of 10 different subjects normally taken. Key Stages 2 and 4 test results receive a lot of attention nationally as they play a prominent role in the computation of so-called school league tables, which are used by policymakers to assess schools and by parents to inform school choice.

3.2 | Outcome and observed background

We focus on GCSEs (Key Stage 4) because they mark the first major branching point in a young person's educational career, and lower levels of GCSE attainment are likely to have a longer term impact on experiences in the labor market. We focus our analysis on results in English and mathematics, which are directly comparable to test results at the end of primary school.[9] Students receive a grade for each GCSE course, where pass grades include A*, A, B, C, D, E, F, G. We use a scoring system developed by the Qualifications and Curriculum Authority to transform these grades into a continuous point score, which we refer to as the Key Stage 4 score.[10]

We control for lagged cognitive achievement using Key Stage 2 National Curriculum tests taken at the end of primary school, usually at age 11, in English and mathematics. In the Key Stage 2 exams, pupils can usually attain a maximum of 36 points in each subject, but teachers will provide opportunities for very bright pupils to test to higher levels. All test scores are standardized separately by subject to have a mean of zero and a standard deviation of one.[11]

The NPD annual school census provides a number of individual and family background variables. These include month and year of birth and gender of the student, ethnicity, whether or not the first language spoken at home is English, any special educational needs identified for the child, eligibility for free school meals (FSM),[12] area of residence and the number of siblings in the family. Because we control for child fixed effects in all our models we do not use these variables as explanatory variables, apart from an interaction term between pupil gender and subject-specific effects to control for gender differences in attainment. We do use several of the background characteristics in the estimation of heterogeneous spillover effects by subgroups and in our robustness checks.

[9]Science exams at age 16 can take different formats, with pupils choosing between single, double, and triple science where results are not necessarily comparable.

[10]A pass grade G receives 16 points, and 6 points are added for each unit improvement from grade G. We use the final GCSE grade attained by each student, which is the grade attained after possible resits. We find that results remain the same when instead using first grade obtained for students taking GCSE exams 1 year early.

[11]The child fixed effect takes into account the variation in the mean of test scores across cohorts, and the variance across cohorts hardly varies.

[12]FSM eligibility is linked to parents' receipt of means-tested benefits such as income support and income-based job seeker's allowance and has been used in many studies as a low-income marker (see; Hobbs & Vignoles, 2010, for some shortcomings).

3.3 | Sibling definition

The NPD includes address data, released under special conditions, which allows us to match siblings in the data set in the year 2007. Siblings are therefore defined as pupils in state schools aged 4–16 and living together at the same address in January 2007. Siblings that are not school age, those in independent schools, and those living at different addresses in January 2007 are excluded from our sibling definition. Step- and half-siblings are included if they live at the same address, and we are not able to distinguish them from biological siblings (see Nicoletti & Rabe, 2013, for details).

3.4 | Peer ability

For each older sibling in our data set we construct a measure of peer ability based on the peers' end-of-primary school test scores that are unaffected by the older sibling. This is our instrumental variable for the older sibling's test score. By using information on the primary schools attended by all pupils we restrict this measure to the new peers encountered by the older sibling for the first time in secondary school. Each student in our sample has 195 cohort peers on average, of which 168 (86%) are new peers. The distribution of the share of new peers is skewed to the left; at the 25th percentile of the distribution 80% of the older sibling's peers are new peers and at the 75th percentile 96% are. As class identifiers are not available in the data we use grade-level ability to proxy the quality of peers experienced by the older sibling. We follow Gibbons and Telhaj (2016) and Lavy et al. (2012) in expressing peer ability in terms of percentiles by subject.[13] Because measurement errors in the older sibling's new peers' achievements at 11 are unlikely to be correlated with measurement errors in either the older or younger sibling's test score at 16, these measurement errors in peer quality may bias the effect of the older sibling's new peers downward in the first-stage estimation but should not bias our instrumental variable estimates.

3.5 | Sample restrictions

The main sample for our analysis includes all sibling pairs taking their Key Stage 4 exams in 2007, 2008, 2009, or 2010. We remove from the data all twins and siblings attending the same academic year, which means siblings can be spaced from one to three academic years apart. When we have multiple pairs of siblings from one family in the observation window we consider the two oldest students to avoid any multiplier spillover effects (what; Dahl, Løken, & Mogstad, 2014, call the snowball effect).[14] We also remove pupils with duplicate data entries or with missing data on background variables from the data set and retain only pupils for whom we have nonmissing test scores for English and mathematics at both Key Stages 2 and 4. This leads to a reduction in sample size of 6.2%. Missing cases are concentrated among low-attaining students that are more likely to be absent at the exams or, at Key Stage 4 are for some reason not entered into their mathematics and/or English examination. Comparing the original with the retained sample the average test score is increased by about 0.7%. We also exclude “special schools" that exclusively cater for children with specific needs—for example, because of physical disabilities or learning difficulties—, as well as schools specifically for children with emotional and/or behavioral difficulties, but retain middle and “all through” schools. Further, we adopt some of the sampling restrictions used in Lavy et al. (2012); namely, we exclude secondary schools with fewer than 15 pupils and schools where the fraction of pupils below the 5th or above the 95th percentile exceeds 20%. The final sample contains 466,392 siblings (233,196 sibling pairs) who go to 2,966 secondary schools in England. We use data that is pooled across two subjects, so that we have 466,392 sibling pair observations in total.

3.6 | Descriptives

Table 1 reports the means and standard deviations of the unstandardized test scores at age 11 and 16 (Key Stages 2 and 4) respectively; but in all our estimated models we consider the standardized test scores by subject. The bottom panel of the table provides proportions (in %) of other characteristics used for the estimation of heterogeneous spillover effects and in our robustness analysis.

[13]We have also experimented with expressing peer ability in terms of standard deviations and with expressing both siblings' outcomes and older sibling's peer ability in percentiles. The results do not change when using different definitions of instruments and outcomes, but the specification we use is the one which gives us the largest *F*-test of the first-stage regression in the 2SLS estimation.

[14]The percentage of siblings who have more than one older siblings is about 1.4% of the sample which already excludes twins and siblings attending the same academic year.

TABLE 1 Descriptive statistics

	Older		Younger	
	Mean	**SD**	**Mean**	**SD**
Unstandardized test scores				
Key Stage 2 English score (age 11)	26.87	4.65	26.12	5.13
Key Stage 2 maths score (age 11)	27.25	4.99	26.94	5.40
Key Stage 4 English score (age 16)	40.00	9.41	39.54	9.14
Key Stage 4 maths score (age 16)	38.99	10.89	38.73	10.63
Sibling characteristics	%			
Same school	84.6			
Brothers	25.8			
Older brother, younger sister	24.8			
Older sister, younger brother	24.3			
Sisters	25.1			
Age gap 1 year	29.3			
Age gap 2 years	49.3			
Age gap 3 years	21.4			
2 children in family	59.0			
3+ children in family	41.0			
Free school meal eligible	11.7			
English additional language	8.5			
No. of observations pooled across subjects	466,392			
No. of sibling pairs	233,196			
No. of schools	2,966			

Note. National Pupil Database, 2007–2010.

TABLE 2 Identifying variation in test scores and instrumental variable

	Mean	**SD**
Younger sibling's test scores at 16		
Total variation	0.020	0.945
Variation net of child fixed effect	0.000	0.341
Variation net of child and school–cohort–subject fixed effects	0.000	0.323
Instrumental variable: KS2 percentiles		
Total variation	49.465	9.596
Variation net of child fixed effect	0.000	2.068
Variation net of child and school–cohort–subject fixed effects	0.000	1.582
No. of observations	466,392	

Note. National Pupil Database, 2007–2010. The instrumental variable is the average of the subject-specific Key Stage 2 test score percentiles across the older sibling's new school peers, excluding the older sibling.

Table 2 gives an overview of the identifying variation in our dependent variable—that is, the younger sibling's standardized school test score at age 16—and in our instrumental variable, which is the average of the subject-specific Key Stage 2 test score percentile across the older sibling's new school peers. The top panel of Table 2 shows the mean and total variation measured by the standard deviation of the younger sibling's test score, the variation net of the child fixed effect, and finally the variation net of both the child and of the school-by-cohort-by-subject fixed effects (see first, second, and third rows). The within-child variation is about 12% of the overall variation. Further applying a child and school-by-cohort-by-subject fixed-effects estimation does not reduce the variation in the data by much.

The bottom panel of Table 2 shows the variation in our instrument. First, we show the total variation in the mean test score percentiles of the older siblings' new peers, excluding the older sibling. On average, the older sibling has 168 new

peers in the same school–cohort. The total variation in the average peer test score percentile is 9.60.[15] By considering the variation net of the child fixed effect we capture the extent to which the older sibling's peers are relatively better in one subject than the others—for example, because they have a good teacher in a particular subject. The SD net of the child fixed effect is 2.07. The last row of the table shows the variation in the data net of both the child and school-by-cohort-by-subject fixed effect.

4 | EMPIRICAL RESULTS

4.1 | Main empirical results

We begin by reporting in Table 3 the correlations in siblings' test scores, which are a general measure of the importance of background shared between siblings for educational outcomes. Since the test scores at ages 11 and 16 are standardized by subject to have zero mean and unit variance, we can estimate the raw correlation in test scores by a simple regression of the test scores at age 16 on the sibling's test score at age 16.[16] This produces the so-called sibling intraclass correlation, which does not generally capture a causal peer effect (see; Angrist, 2014). The raw correlation in test scores is shown in column (1) of Table 3 and estimated to be 0.476, which is in line with results of previous papers (e.g., Björklund, Eriksson, & Jäntti, 2010; Nicoletti & Rabe, 2013).

In column (2) we display the sibling correlation in test scores net of the effect of past test scores obtained by the younger sibling at the end of primary school, which we estimate by using a value-added model—that is, by regressing the test scores at 16 on the older sibling's test scores at 16 and controlling for the younger sibling's test scores at age 11 and subject–gender interactions. This value-added model is equivalent to Equation 1 but omits controls for family and school inputs and unobserved individual and school-by-cohort-by-subject characteristics. The sibling correlation now captures the effect of shared family and environment characteristics that operate between ages 11 and 16 and reduces to 0.292. In column (3) we show the correlation estimated using the same value-added model as in column (2) and controlling for the younger child's fixed effects (0.132). This eliminates the influence of all environment, family, and child characteristics that are invariant across subjects, including the intra-household allocation of resources between siblings and the effects of schools of both siblings on the younger sibling. The next step to move from a sibling correlation to a sibling spillover effect is to take account of possible subject-specific characteristics and inputs from school and family.

To control for subject-specific school characteristics and inputs, we estimate the model (Equation 1) controlling for child fixed effects and additionally for school-by-cohort-by-subject fixed effects of the younger sibling (FE estimation).[17] This yields a sibling spillover effect of 0.110, which is statistically significant at the 1% level (see Table 4, first column). This FE estimate comes close to capturing a causal relationship, but it could still be overestimated because of unobserved subject-specific skills transmitted in the family that are similar between siblings. Families are likely to have subject-specific traits—being a "math" family, for example—which can affect both subject-specific inherited child endowments and subject-specific family investments. To check and control for any such residual endogeneity bias in the FE estimation, we estimate Model 1 using instrumental variable estimation (FE estimation with IV). This is a 2SLS estimation where the first stage consists in the regression of the older sibling's test score at age 16 on all explanatory variables plus an instrument given by the average subject-specific ability at age 11 of the older sibling's new school peers encountered for the first time in secondary school; whereas the second stage is the regression of the younger sibling's test score on all explanatory variables and with the older sibling's test score replaced by its prediction from the first-stage regression. Because we control for the younger sibling's individual fixed effect, school-by-cohort-by-subject fixed effect, and subject-specific average test score of the younger sibling's new peers, our instrument captures whether the older sibling's new school–cohort mates were relatively better in a specific subject than the younger sibling's new school–cohort mates—for example, because of changes in the quality and quantity of school inputs in a specific subject or in the subject-specific abilities of the school–cohort mates.

The second column of Table 4 shows the estimated sibling peer effect using the FE estimation with IV, which suggests that an increase of one SD in the test score of the older sibling leads to an increase of 15.6% of a SD in the corresponding

[15]Note that the SD is lower than the SD of a percentile score—that is, the SD of a discrete uniform (1,100), which is 28.86, as we average the percentiles across the older sibling's new peers.
[16]For all regression models we allow the error terms to be clustered at school level and report robust standard errors.
[17]We are not concerned about the endogeneity of the lagged test caused by the fact that child unobserved endowments influence both the test scores at ages 11 and 16 because we control for child fixed effects and therefore eliminate child unobserved endowments. This method to correct for the endogeneity of the lagged test has previously been applied in Nicoletti and Rabe (2018), Slater et al. (2010), and Del Boca, Monfardini, and Nicoletti (2017), for example.

TABLE 3 Sibling correlations in test scores

	(1) Raw correlation	(2) Correlation value added	(3) Correlation value added Child FE
Corr.	0.476**	0.292**	0.132**
	(0.002)	(0.002)	(0.002)
Observations	466,392	466,392	466,392

Note. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$. National Pupil Database, 2007–2010. Standard errors clustered at school level in parentheses. Pooled sample, pooling the observations for mathematics and English. The value-added model in column (2) and the model in column (3) control for younger siblings' age 11 test scores and subject-by-gender dummies. Column (3) includes child fixed effects.

TABLE 4 Sibling spillover effect: Main results

	(1) Child–School–Coh.–Subj. FE without IV	(2) Child–School–Coh.–Subj. FE with IV
γ	0.110**	0.156*
	(0.002)	(0.068)
First-stage coefficient		0.0059**
		(0.001)
F-test first stage		90.29
Stock–Yogo critical value (10% maximal IV size)		16.38
Endogeneity test		0.471
p-value		(0.493)
Observations	466,284	466,284

Note. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$. National Pupil Database, 2007–2010. Child and school-by-cohort-by-subject fixed-effect estimations with and without instrumental variable presented in columns (1) and (2). Dependent variable is the within-child deviations from mean standardized Key Stage 4 scores in English and mathematics. Value-added model controls for age 11 test scores and subject-by-gender dummies. Standard errors clustered at school level in parentheses. Pooled sample, pooling the observations for mathematics and English. The instrument is the deviation from the average Key Stage 2 attainment percentile in English and mathematics of the older sibling's new peers in secondary school. The *F*-test is the Angrist–Pischke multivariate *F*-test of excluded instruments in the first stage. The endogeneity test is the robust Durbin–Wu–Hausman test.

test score of the younger sibling, and this effect is statistically significant at the 5% level. The endogeneity test reported in Table 4 suggests, however, that after controlling for child fixed effects and school-by-cohort-by-subject fixed effects there is no residual endogeneity of the older sibling's test score, and we cannot reject the equality of the FE estimation and FE estimation with IV (first and second columns in Table 4). The estimation with just fixed effects is more precise and therefore our preferred estimation, and we will use it to produce estimates that allow for a heterogeneous sibling spillover effect (see Section 4.3).

The *F*-statistic for the significance of the instrumental variable in the first stage is large and does not leave any doubt on the validity of the instrument. Our first stage coefficient is 0.0059, which means that a 1 SD increase in average peer ability (9.596) increases older sibling's ability by 5.7% of a standard deviation. While this effect size is not small in comparison to educational interventions, we acknowledge that a prediction of the effect of a 1 SD increase in the older sibling's test score on the younger sibling's test score requires an out-of-sample prediction with respect to the range of values taken by our instrument, therefore relying on functional form assumptions; see Carrell, Sacerdote, and West (2013) and Booij, Leuven, and Oosterbeek (2017) on the often overlooked issue of limited support.[18]

[18]The range of values taken by our instrumental variable is 0–97, so that a change from the bottom to the top of the range would increase the older sibling's test score by 60% of a SD.

Our first-stage regression is similar to the model adopted in Lavy et al. (2012) to estimate school peer effects using the same school administrative data that we use. As in Lavy et al., we are concerned with the reverse causality that goes from older siblings to their school peers and deal with this by using predetermined peer ability measures. But, contrary to them, we are not interested in interpreting the coefficient of the ability of the older sibling's school peers as an endogenous school peer effect. Indeed, because we do not control for the school-by-cohort-by-subject fixed effect of the *older* sibling, the average ability of the older sibling can be driven by variation in school inputs which are subject and cohort specific (e.g., quality of math teachers). This implies that our estimated effect of the older sibling's school peers' ability cannot be interpreted as a school peer effect on the older sibling but rather as the combined effect of peers and of school inputs that are subject and cohort specific. This explains why our first-stage effect is about two to three times the size of that estimated by Lavy et al.[19] Note, however, that because we control for the school-by-cohort-by-subject fixed effect of the younger siblings our instrumental variable remains valid even if there is correlation in unobserved subject-specific school inputs between siblings.

Using our estimates of the sibling spillover effect, we can quantify how much the effect of school policies or investments targeted at all children would be amplified by interactions between siblings. For example, consider a school intervention that leads to 10% of a SD increase in test scores for all children. The total effect of the intervention would be 11.1% of a SD, of which 10% of a SD is the direct effect of the intervention and 1.1% of a SD is the indirect effect through the sibling spillover. The indirect effect is computed by multiplying the direct effect of the intervention on the older sibling, 10% of a SD, by the estimated spillover effect on her younger sibling, 0.110.

We can also compare the sibling effects we estimate to those obtained for school peers and school friends in previous papers. School peer effects estimated in recent paper are zero or very small (Del Bello, Patacchini, & Zenou, 2015; Gibbons & Telhaj, 2016; Lavy et al., 2012), whereas effects based on nominated school friends are higher at around 10% of a SD (Del Bello et al., 2015; Patacchini, Rainone, & Zenou, 2011). This seems to suggest that sibling interactions are comparable to interactions between school friends and more relevant than interactions between school peers.

4.2 | Threats to identification: Robustness checks

In this section we discuss threats to the validity of our identification strategy and probe the stability of our 2SLS estimates to robustness checks and alternative specifications.

We start off by showing results of IV regressions that omit the child fixed effect. In principle, if the IV strategy is valid, the child fixed effect should not be needed for identification as it would account for the problem of omitted family inputs. As can be seen in the first panel of Table 5, the estimated spillover effect is slightly larger when omitting child fixed effects (0.183 vs. 0.156) but the difference is not sizable or statistically significant. These results provide evidence that after controlling for child fixed effects there is still enough variation in the child's test scores and in the instrumental variable to identify the sibling spillover effect. The IV estimation without child fixed effects has the advantage that it is not taking out any general skill that may be transmitted from older to younger siblings, but it also does not control for unobserved child characteristics that could be correlated with both the test score at 16 and the lagged test score at 11. For this reason we prefer to use the child fixed-effect estimation.

4.2.1 | Weak instrument and falsification tests

We test whether, despite the high F-statistic of 90.29 of the first stage, we have a weak instrument problem in our IV estimation. First, we run a Hahn and Hausman (2002) test of a strong instrument, which involves running a reverse IV regression of older on younger siblings' test scores, using the older siblings' new peers as instruments for younger siblings' outcomes. If the instrument is strong, the regression and reverse regression should give estimates of γ and $1/\gamma$ that are consistent with each other. The top row in the second panel of Table 5 shows the result of this exercise. Reassuringly, the reverse regression provides an estimate for $1/\gamma$ of 6.406 and its reciprocal is 0.156, which is identical to our instrumental variable estimate of γ (see Table 4).

Next we perform a falsification test, where we replace each older sibling's true peers with similar, but randomly drawn, peers. We want to make sure that there is no effect of randomly drawn peers on the older sibling's attainment in the first stage and no spillover effect estimated in the second stage. Specifically, we assign to each older sibling the peer group of a

[19]Further differences between our first-stage estimation and the main equation estimated in Lavy et al. (2012) include (i) the estimation method (we control for younger sibling's school-by-cohort-by subject fixed effects in addition to the child fixed effect), (ii) the point in time when school test scores are measured (at 16 rather than 14), and (iii) the selection of the sample (e.g., we do not focus on small schools).

TABLE 5 Robustness checks

(1) Sibling spillover	(2) *F*-test	(3) Endogeneity test	(4) Further tests
IV without child fixed effect			
0.183**	4,204	37.56	
(0.009)		(0.000)	
Weak instrument and falsification tests			
Hahn–Hausman (2002) weak instrument test (reverse IV)			1/coeff.
6.406*	5.254		0.156
(2.770)			
Falsification test: assigning randomly drawn peers to older sibling			Times reject zero
0.493	0.997		0.3%
(465.88)			
Falsification test: assigning a similar older child to be older sibling			Times reject zero
−0.002	327.6		4.4%
(0.049)			
Falsification test: assigning child from older sib.'s school to be older sib.			Times reject zero
−0.001	147.2		4.6%
(0.049)			
Reverse regression from older to younger			
−0.008	93.33	2.755	
(0.067)		(0.097)	
Tests for possible interaction with older sibling's peers			
Excluding older sibling's school mates living in the same neighborhood			
0.156*	88.73	0.458	
(0.068)		(0.498)	
Excluding older sibling's schoolmates living in the same area			
0.148*	94.08	0.318	
(0.068)		(0.573)	
Testing overidentification			
1. New peers' KS2 percentiles and best KS2 subject			Hansen's J
0.166*	60.21	2.097	1.944
(0.066)		(0.148)	(0.163)
2. New peers' KS2 percentiles and percentage of KS2 bottom 5% pupils			
0.154*	45.00	0.490	1.108
(0.067)		(0.484)	(0.292)

Note. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$. National Pupil Database, 2007–2010. Standard errors clustered at school level and *p*-values in parentheses. Pooled sample, pooling the observations for mathematics and English. The *F*-test is the Angrist–Pischke multivariate *F*-test of excluded instruments in the first stage. The endogeneity test is the robust Durbin–Wu–Hausman test. See text for description of the tests and checks. "Neighborhood" refers to the Lower Level Super Output Area, "Area" to the Middle Layer Output Area of residence. Additional instruments are the proportion of new peers that had English or mathematics as their best subject at the end of primary school and who were in the bottom 5% of pupils in Key Stage 2.

randomly drawn older child, where the older child has the same gender, ethnicity (white British vs. not), free school lunch eligibility, and language spoken at home (English vs. not) as the true older sibling. We run this procedure 1,000 times and summarize the results in the second panel of Table 5. We report the mean of the estimated sibling spillover effect across the 1,000 replications and between parentheses the corresponding mean of the standard error. The mean *F*-statistic of the first stage is very small (0.997) and the null of a zero spillover effect is rejected in 0.3% of the iterations, indicating that our results are not replicated using random peers.

We carry out a further falsification test similar to the previous one, but this time assigning an unrelated child to be the older sibling while we keep using the true school mates of the assigned unrelated child to derive the instrumental variable. Here we expect the first stage to work, as each older child's ability would be instrumented by her true peers, but we do not expect a spillover effect because the fictitious siblings we have matched do not in fact interact with each other. As before, we assign a child as older sibling who shares the true sibling's gender, ethnicity, free school meal status, and language

spoken at home, and perform this draw 1,000 times. As can be seen from Table 5, the first stage has a high *F*-statistic and the spillover effect estimate is zero. The null of a zero effect is rejected in less than 5% of the 1,000 replications, confirming that no spillover effect is estimated from older to unrelated younger children.

Our third falsification test addresses the concern of endogenous school choice, in particular the concern that variation in the older sibling's schoolmates' performance arising from older siblings attending different schools might be endogenous. To test whether endogenous variation is driving our results we swap real older siblings with randomly chosen children from the older sibling's school 1,000 times. Table 5 shows that the mean first stage from this procedure is high, as expected, and that there is no spillover from the randomly chosen sibling. We reject a zero spillover in 4.6% of cases.

We have also performed further checks around the role of school choice for our results, including restricting the estimation sample to sibling pairs from the same primary and/or secondary school; and within older siblings' schools to a group of older siblings who shared the same primary school. While slicing the data in this way weakens the first stage and drives up standard errors of the sibling spillover effect, the *F*-test in the first stage is still large enough for the instrument to pass weak instrument testing and the endogeneity test still indicates that we cannot reject exogeneity of the older sibling's test score. Furthermore, the preferred fixed-effects estimation still provides results in line with our benchmark results (available on request).

The final check in the second panel of Table 5 is the reverse IV regression from the younger to the older sibling. The model specification is identical to Model 1, with the subscripts 1 and 2 exchanged to swap the role of the younger sibling with that of the older sibling and using the younger sibling's new peers' average attainment in primary school as instrument. As we can see, the first stage is quite strong and the second stage shows no spillover effect from the younger to the older sibling, indicating that there is no reverse causality at play. Note that this spillover effect is not directly comparable to that running from the older to the younger sibling, because (i) the younger sibling is exposed to new peers for a much shorter time until the older sibling's attainment is measured (1–3 years) than vice versa (5–7 years), and (ii) the younger sibling's test score at 16 is observed in the future with respect to the older sibling's test score at 16.

4.2.2 | Direct influence of older sibling's schoolmates on the younger sibling

One of our identifying assumptions is that the older sibling's peers have no direct influence on the younger sibling's test scores. We investigate here the possibility that the older sibling's schoolmates could directly interact with the younger sibling in the neighborhood or at school and therefore violate the exogeneity assumption.

Although peers from the older sibling's primary school, including "forever friends" whom the younger sibling may know and have interacted with as a child, are excluded from the instrument, it may be that some new secondary school peers live in the same neighborhood and interact with the younger sibling even if they do not belong to the same cohort. Evidence for England shows that there are no neighborhood peer effects in school achievement (Gibbons, Silva, & Weinhardt, 2013), but we still want to test this possibility. In our data, we can define neighborhoods based on Lower Level Super Output Areas, which are statistical geographies created to reflect proximity and social homogeneity that have an average of roughly 1,500 residents in 650 households. In our sample, an average of nine peers from the same school and cohort live in a neighborhood defined in this way (a school cohort comprises 195 pupils, on average). Among these, on average five students are old and four are new peers. Secondary students may interact within a wider geographical area, so we also look at Middle Layer Output Areas (which, on average, have a population of 7,500 and 3,000 households). An average of 36 peers from the same school and cohort live in an area thus defined, of which 23 are new peers. We take this as the maximum proportion of the older sibling's schoolmates a (very sociable) younger sibling could be exposed to within the residential area.[20]

To test the possibility of neighborhood interaction, we exclude the older sibling's new school peers living in the same neighborhood (and alternatively the wider area) from the computation of the instrumental variable to remove the potential direct effects that go from children living in the same neighborhood (area) to the younger sibling. Table 5 displays the results of this exercise. Excluding the older sibling's new schoolmates living in the same neighborhood from the calculation of the instrument changes the estimated sibling spillover effect by very little. Excluding the older sibling's schoolmates living in the same area again produces a result that is comparable to the benchmark estimate. This suggests that direct interaction within neighborhoods and wider areas does not threaten our identifying assumption.

Another possibility is that younger siblings directly interact with their older sibling's schoolmates at school. However, unlike the cases of Bramoullé et al. (2009) and Calvó-Armengolo, Patacchini, and Zenou (2009), where the unrelated peers

[20]Note that Middle Layer Output Areas are quite large geographical areas, with an average size of 1,958 hectares across England (1 hectare = 10,000 m^2), which cannot easily be traversed by a child on a regular basis, in particular in rural areas.

of peers can be taught in the same school class, in our case the older sibling's peers are in different classes and cohorts from the younger sibling, sometimes several years apart. In English schools, cohorts are taught strictly separately, and because of the large cohort size of secondary schools even school assemblies usually take place separately by cohort. This means that interactions that are relevant for learning are unlikely to take place across cohorts in school. Unfortunately we have no way of testing this further.

4.2.3 | Exploring additional instruments

Next we check the validity of our instrument further by using additional instruments, which allows us to test the overidentifying restrictions. We consider as first additional instrument the proportion of the older sibling's schoolmates that had a particular subject as their best subject. This may reflect the selection of similarly talented students into the same school or the presence of better teachers in a specific subject within a school. As we can see in the first row of the bottom panel of Table 5, the *F*-test of the excluded instruments is very high, indicating that the instruments are relevant, and the estimated sibling spillover effect remains the same as before. Hansen's J test shows that the null that the instruments are exogenous cannot be rejected.

We consider as our second additional instrumental variable the fraction of the older sibling's new peers that were in the bottom 5th percentile of the subject ability distribution at the end of primary school. This variable is identical to the one used in Lavy et al. (2012) to estimate school peer effects. We show the results of our IV estimates using both our original instrument and the fraction of bad peers of the older sibling as instruments in the second row of the bottom panel of Table 5. As we can see, this does not change the results and Hansen's J test suggests that our instruments are valid.

4.2.4 | Further threats and checks

Our instrumental variable could fail because of the way our sample is constructed. We have data for four cohorts of students taking age 16 exams, and it is possible that an older sibling has schoolmates whose younger siblings are in the same cohort and same school as her younger sibling. In this case there could be a direct effect of the older sibling's schoolmates on the younger sibling through their younger siblings. However, because we control for the younger sibling's school-by-cohort-by-subject fixed effects, any link to the older sibling's schoolmates through the younger siblings' schoolmates is broken.

Our instrument could also fail because of possible sorting of both the younger and older sibling into schools with similar peers and characteristics; but, as emphasized in Section 2, the school–by–cohort–by–subject fixed effect controls for these similarities and our instrumental variable approach exploits only idiosyncratic variation in the composition of school peers across cohorts within a school and/or between schools. Moreover, we control for the end of primary school test scores of the younger sibling's peers in our estimates, alleviating such concerns further.

We carry out two further checks for which we do not report the results (they are available on request). First, our identification strategy relies on the assumption that subject-specific test scores follow the same production model as general cognitive ability. We test this by checking that spillover effects are the same across both subjects (math and English) and confirm that they are.

We also test whether estimated effects are the same when limiting the sample to small schools with below median cohort size. It could be that students in small, rural schools do not meet many new peers and results are weighted towards larger schools. Both fixed effects and IV estimation for small schools gives results that are identical to our baseline results shown in column (4) of Table 4 and in Table 5; the FE point estimates are 0.106 and the IV point estimates are 0.139.

4.3 | Heterogeneity and implications for inequality

In this section we perform subgroup analysis to explore what we can learn about possible mechanisms that may drive the sibling spillover effects (e.g., imitation, productivity spillovers, and information transmission) and to consider the implications of our results for inequality between students from different socioeconomic backgrounds.

We begin by estimating spillover effects from the older to the younger sibling by sex composition and age gap between the siblings (measured in academic years), as well as family size and school attended. If both siblings attend a school with an attached "sixth form" (post-compulsory school catering for students aged 17 and 18), then the older sibling may remain at the same school after their school-leaving exams at age 16, making interaction with the younger sibling remaining at

TABLE 6 Heterogeneous sibling spillover effects

	(1)	(2)	(3)	(4)
Sex combination	Brother→brother	Brother→sister	Sister→brother	Sister→sister
	0.120**	0.097**	0.104**	0.122**
	(0.004)	(0.004)	(0.004)	(0.004)
Age gap	1 year	2 years	3 years	
	0.111**	0.112**	0.106**	
	(0.004)	(0.003)	(0.004)	
Family size	2 siblings	3+ siblings		
	0.110**	0.111**		
	(0.003)	(0.003)		
School	Different	Same, no 6th form	Same, 6th form	
	0.091**	0.112**	0.116**	
	(0.005)	(0.004)	(0.003)	
Older sib. KS4 results	Bottom grades	Medium grades	Top grades	
	0.068**	0.122**	0.161**	
	(0.003)	(0.002)	(0.004)	
Observations	466,392			

Note. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$. National Pupil Database, 2007–2010. Pooled sample, pooling the observations for mathematics and English. Results are from child and school-by-cohort-by-subject fixed-effect estimation with interaction terms used to derive coefficients by subgroup within each panel. Standard errors clustered at school level in parentheses. Age gap is measured in academic years. Older siblings attainment at Key Stage 4 grouped by grade, where top grades are A* and A, medium grades B–D, bad grades E–G and U.

the school more likely. Results are shown in the first three panels of Table 6.[21] We might expect siblings who are of the same sex, closer in age, from smaller families and in the same school to interact more and feel closer to each other and therefore to be more likely to interact through imitation, direct help/teaching or information sharing. We do find that the sibling spillover effect on the younger child's test score is about 2% of a SD higher for siblings of the same gender (brother and sister pairs) than for mixed-gender siblings and for siblings in the same school, and somewhat larger for siblings who are more closely spaced. We do not find spillover effects to be larger between siblings who have no further siblings to interact with.

Next we split the sample by the older sibling's attainment. We distinguish top attainment, defined as having an A* or A grade in a subject (16% of the sample), bottom attainment (grades E–G and U, 18% of the sample) and medium attainment (grades B–D, 65% of the sample). As the bottom panel of Table 6 shows, here we find substantial differences in the spillover effects, with spillovers from top-achieving older siblings being more than twice as high as those from bottom-achieving ones (16.1% vs. 6.8% of a SD). Spillover effects from the middle group are in the region of the average spillover effects we find: 12.2% of a SD. This gradient in the spillover effect across the older sibling's achievement could suggest that productivity spillovers are quite important. These are produced through learning of the younger sibling from their older sibling, for example by spending time together in doing formative activities, by being taught, or by receiving help with their homework, and should arguably be larger when the older sibling performs well at school as this will affect the quality of the interaction. Alternatively we can speculate that the older sibling becomes a more inspiring role model once he or she is a top attainer; that is, there may be nonlinearities in role model effects.

To explore the gradient of the spillovers across older siblings' attainment further, we examine effects by family background. We might expect that children from disadvantaged families have parents who are less likely or able to help them in their learning and that sibling interactions play a larger role in such families. We measure family disadvantage in three different ways: by deprivation of neighborhood of residence,[22] eligibility for free school lunches, and by whether the language spoken at home is English. Neighborhood deprivation captures income deprivation of the area, while free school meal eligibility indicates low income in the student's household. Limited language proficiency is a further dimension of disadvantage.

[21] All heterogeneous sibling spillover effects are estimated by interacting the older sibling's subject-specific test score with dummy variables for different subgroups and using our preferred estimation, which is the estimation with child and school-by-cohort-by-subject fixed effects.
[22] Deprivation is measured using the Income Deprivation Affecting Children Index at the Lower Level Super Output Area, which is a subdomain of the English Indices of Deprivation. We divide children's neighborhoods into the most, middle, and least deprived tertiles.

TABLE 7 Sibling spillovers by family background

	(1)	(2)	(3)
Free school meal status	**FSM eligible**	**Not FSM eligible**	
Older sib. bottom grade	0.066**	0.069**	
	(0.008)	(0.004)	
Older sib. medium grade	0.096**	0.126**	
	(0.007)	(0.002)	
Older sib. top grade	0.168**	0.161**	
	(0.016)	(0.004)	
All	0.084**	0.115**	
	(0.006)	(0.002)	
Neighborhood deprivation	**Most deprived tertile**	**Middle tertile**	**Least deprived tertile**
Older sib. bottom grade	0.068**	0.068**	0.069**
	(0.005)	(0.006)	(0.007)
Older sib. medium grade	0.114**	0.127**	0.126**
	(0.004)	(0.004)	(0.004)
Older sib. top grade	0.154**	0.166**	0.161**
	(0.008)	(0.006)	(0.005)
All	0.097**	0.115**	0.123**
	(0.003)	(0.003)	(0.003)
Language at home	**Not English**	**English**	
Older sib. bottom grade	0.094**	0.066**	
	(0.011)	(0.004)	
Older sib. medium grade	0.165**	0.118**	
	(0.008)	(0.002)	
Older sib. top grade	0.200**	0.158**	
	(0.012)	(0.004)	
All	0.148**	0.107**	
	(0.007)	(0.002)	
Observations	466,392		

Note. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$. National Pupil Database, 2007–2010. Pooled sample, pooling the observations for mathematics and English. Results are from child and school-by-cohort-by-subject fixed-effect estimation. Standard errors clustered at school level in parentheses. Each panel includes results from two regressions, one with interaction terms capturing family disadvantage, the other using disadvantage × older sibling attainment interaction terms. Older siblings attainment at Key Stage 4 grouped by grade, where top grades are A* and A, medium grades B–D, bad grades E–G and U.

Table 7 shows results by free school meal status, neighborhood deprivation, and language spoken at home in separate panels. Within each panel the first three rows give results from models where family background is interacted with older siblings' attainment. The differences in spillovers between children with top-, medium-, and bottom-attaining older siblings are large, but within the attainment groups they do not vary much by family background. For example, the spillover effect of a bottom-attaining older sibling is nearly identical across all types of neighborhood deprivation and across free lunch status, and the same holds true for a top-attaining older sibling.[23] This is a positive and encouraging result; not only is the sibling peer effect higher for children whose older siblings are performing well rather than badly in school, but also this does not differ (much) by family background.

Within each panel of Table 7 the last row, labeled "All," gives results by family background without older siblings' attainment interactions. Looking just at family background changes the picture somewhat. We see that the overall sibling spillover effect is lower for children who are eligible for free school lunches or live in deprived areas than for children from more affluent backgrounds. This is because the overall effect is essentially a weighted average of the effects by attainment of the older sibling, where the weights are the proportions of older sibling in each attainment group. Low-income students eligible for free lunches are more likely than high-income students to have low-attaining older siblings (36% vs. 16% of

[23]Spillover effects among students who speak English as their first language are higher than those of students who do not, but the differences are relatively stable across attainment of the older sibling.

TABLE 8 Average effect of sibling spillovers by family background

	(1)	(2)	(3)	(4)
	FSM eligible	**Not FSM eligible**		**(1)–(2)**
Av. effect	−0.015**	0.036**		−0.051**
	(0.000)	(0.000)		(0.000)
	Most deprived nbh	**Middle**	**Least deprived nbh**	**(1)–(3)**
Av. effect	−0.000**	0.032**	0.058**	−0.058**
	(0.000)	(0.000)	(0.000)	(0.000)
	Not English	**English**		**(1)–(2)**
Av. effect	0.025**	0.030**		−0.005**
	(0.000)	(0.000)		(0.000)

Note. $^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$. National Pupil Database, 2007–2010. The average effect is the spillover effect by attainment and family background from Table 7, multiplied by the average standardized point score in the attainment category and the proportion of students whose older sibling is in that attainment category, summed over the three attainment categories. nbh, neighborhood.

students) and considerably less likely to have top-attaining older siblings (5% vs. 18% of students), leading to an overall lower spillover effect for this group.[24] Taken together, our results indicate that the lower sibling spillover effects found in low-income families are not caused by the fact that interactions in these families differ from those in higher income families, but by the fact that children in these families are more likely to have a low-attaining older sibling.

The fact that the spillover effect is smaller for children from disadvantaged than from affluent backgrounds has implications on the effect of school policies that aim at raising school achievement for all children. This is because the direct policy effect will be amplified by an indirect effect through interactions between siblings that is larger for children from privileged backgrounds and therefore leads to an increase in the attainment gap between children from disadvantaged and affluent backgrounds. A school intervention that increases test scores of all children by 10% of a SD would lead to a total effect of 0.84% of a SD for FSM-eligible children and 1.15% for non-FSM-eligible children (compare first and second columns in the last row of the top panel of Table 7) and therefore to an increase in the gap of 0.3% of a SD, assuming that there is nobody moving across bottom, medium, and top grades because of the reform. Similarly, this type of intervention would increase the gap between children who live in more or less deprived areas by 0.3% of a SD (compare columns (1) and (3) in the last row of the central panel of Table 7). Conversely, the same intervention would reduce the gap between children who speak and do not speak English at home by 0.4% of a SD (compare columns (1) and (3) in the last row of the bottom panel of Table 7).

To take these results one step further we next derive the average impacts on test scores of having an older sibling, by family background. This measures the expected SD change in the younger sibling's test score for a change in the older sibling's test score from zero (no sibling present) to the average test score observed for older siblings in a family with given background characteristics. In other words, we ask what happens to the attainment of a child when placing her in different types of families. For example, the total average impact of having an older sibling who is eligible for free school meals is equal to the weighted average of the impacts across the three attainment categories of the older sibling with weights given by the proportion of students on free lunches with older siblings in the bottom, medium, and top grades respectively.[25] Table 8 shows average effects by free school meal status, neighborhood deprivation, and language spoken at home. Strikingly, being born into a low-income family reduces students' attainment by 1.5% of a SD, on average, through sibling spillovers, whereas being born into a higher income family increases students' attainment by 3.6% of a SD. In total, the difference between the groups is 5.1% of a SD, which amounts to 8.4% of the attainment gap between FSM and non-FSM students (61% of a SD). Similarly, the difference between children living in the least and most deprived neighborhoods is 5.8% of a SD, whereas average test score impacts do not differ much by language spoken at home.

[24]For children who do not speak English at home, the influence of the older sibling is considerably larger than in English-speaking families. It might be that siblings are compensating for the fact that parents may lack knowledge of the English education system as they will in most cases not have been raised and educated in England.

[25]More specifically, we compute the average effect for children on free lunches with a top (medium, bottom)-attaining older sibling by multiplying the estimated spillover effect (0.168, 0.096, 0.066 for top-, medium-, bottom-attaining older siblings from Table 7) by the average standardized point score observed for older siblings who are top (medium, bottom) attainers (1.40 SD, 0.13 SD, −1.46 SD). The resulting average effects (0.235, 0.012, −0.096) are then weighted by the proportion of top (medium, bottom)-attaining older siblings among families eligible for free school meals (0.05, 0.58, 0.36) and summed over the attainment categories to give the average effect (−0.015 of a SD). The effects for the other groups are computed in the same way.

Our results have interesting policy implications. They suggest that policymakers aiming at reducing socioeconomic inequalities between children need to be less worried about the quality of interactions siblings have within families, as spillover effects by attainment of the older sibling do not differ much by background and generally favor high attainment (see Table 7). Instead, the negative impact of older siblings in free school meal eligible families (see Table 8) is, on average, driven by their poor attainment. This means that rather than promoting family-based interventions aimed at changing the interactions between siblings (i.e., the coefficients) policymakers might want to look at interventions to increase the attainment of (disadvantaged) older siblings. However, our analysis also shows that interventions that improve the attainment of all children, regardless of family background, are likely to increase the attainment gaps between children from lower and higher income backgrounds (but close those between children by language spoken at home). This is because the spillover effect is higher for higher income (and non-English-speaking) families, implying a higher indirect effect of any policy for this group. This suggests there is a trade-off between equity and efficiency considerations, with policies aimed at closing the SES attainment gap coming at a higher cost than indiscriminate policies.

5 | CONCLUSIONS

In this paper we provide empirical evidence of sibling spillover effects in school achievement based on administrative data of 230,000 siblings taking their end-of-compulsory schooling (age 16) exams in a 4-year time window. We measure school achievement using test scores obtained in national exams in England in the compulsory subjects English and mathematics. We find strong evidence of sibling spillover effects in school achievement. An increase in the test scores of an older sibling of one SD leads to an increase in the corresponding test score of the younger sibling of about 11% of a SD. In terms of one grade improvement (e.g., from a grade B to an A) this effect is equivalent to the effect of increasing school expenditure per pupil by about £1,000 (see; Nicoletti & Rabe, 2018).

While the previous literature on social networks has provided evidence that children's school achievements are affected by peers outside the household, we show that they are also affected by older siblings. The spillover effect from the older to the younger sibling is comparable in size to the effect of friends and larger than the effect of schoolmates.

Our estimation strategy takes account of the three well-known identification issues in peer effect estimation (Brock & Durlauf, 2001; Dolton, 2017; Manski, 1993; Moffitt, 2001), which are issues of correlated omitted variables, reflection (simultaneity), and endogenous peer membership. We control for these issues by controlling for child and school-by-cohort-by-subject fixed effects and instrumenting the older sibling's test score with his or her peers' predetermined achievement.

The large sample size available in our data allows us to perform subgroup analysis and to explore implications of sibling spillover effects for inequality. While results do not differ hugely by sibling sex combination, age gap, family size, or by whether siblings share the same school, we find striking differences by older siblings' attainment. Spillovers from older siblings with top grades are substantially larger, whereas the impact of badly performing older siblings is considerably smaller on average. This seems to suggest that older sibling are effective teachers or important role models for their younger siblings especially when they perform well in school, while there is some resilience against the impact of badly performing older siblings. We find that, perhaps surprisingly, this pattern holds for children regardless of socioeconomic background. In other words, children from low-income families are helped just as much by their top-attaining older siblings as children from higher income families, and they do not suffer a larger influence from bad performance than their more affluent peers. However, low-income children are more likely to have an older sibling who is not performing well in school. Therefore, on average, the performance of children in low-income families in England decreases as a result of the interaction with older siblings, whereas that of children in higher income families improves. Comparing children eligible and not eligible for free school lunches, we find a gap in the average school attainment score of 61% of a SD and, on average, 8.4% of this gap is explained by the sibling spillover effect.

Taken together, our paper has important implications for policy that seeks to narrow the attainment gaps between children from different socioeconomic backgrounds. Our results indicate that these gaps are not driven by discernible differences in the way that siblings interact within families that might have motivated family-based interventions. Instead, it is the poorer performance of low-income children that transmits to siblings, suggesting that school-based investments into this group are likely to have non-negligible externalities through their benefits for younger siblings. After the time period covered by our analysis, in 2011, the UK government introduced the Pupil Premium in England, which is an example of such a policy. The policy allocates additional funding to schools for each student eligible for free school meals (currently about £1,000 per year in secondary and £1,300 in primary schools), which has to be spent to raise attainments of

these disadvantaged students irrespective of their abilities. If these funds are successful in raising attainment our analysis suggests that spillover effects would work as a multiplier and the socioeconomic attainment gap would narrow. However, investing the same funding into higher income children would likely lead to a higher overall effect, based on higher social multipliers for this group, albeit widening the attainment gap.

ACKNOWLEDGMENTS

We thank the Department for Education for making available, under terms and conditions, data from the National Pupil Database. We thank the editor, three anonymous referees as well as seminar participants at the Labour and Public Policy Seminar at the University of Aarhus, at the Collegio Carlo Alberto and at the Luxembourg Institute of Socio-Economic Research for comments and feedback. This research was supported by core funding of the Research Centre on Micro-Social Change at the Institute for Social and Economic Research by the University of Essex and the UK Economic and Social Research Council (awards RES-518-28-001 and ES L009/53/1).

REFERENCES

Angrist, J. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108.

Björklund, A., Eriksson, K. H., & Jäntti, M. (2010). IQ and family background: Are associations strong or weak? *BE Journal of Economic Analysis and Policy, (Contributions)*, *10*(1), Article 2.

Björklund, A., & Salvanes, K. G. (2011). Education and family background: Mechanisms and policies. In Hanushek, E. A., Machin, S., & Woessmann, L. (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 201–247). Amsterdam, Netherlands: North-Holland.

Booij, A. S., Leuven, E., & Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *Review of Economic Studies*, *84*, 547–578.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*, 41–55.

Breining, S. (2014). The presence of ADHD: Spillovers between siblings. *Economics Letters*, *124*, 469–473.

Breining, S., Daysal, N. M., Simonsen, M., & Trandafir, M. (2015). Spillover effects of early-life medical interventions. (*IZA Discussion Paper No.9086*). Bonn, Germany: Institute of Labor Economics.

Brock, W., & Durlauf, S. N. (2001). Interactions-based models. In Heckman, J. J., & Learner, E. (Eds.), *Handbook of econometrics*, Vol. *5*. Elsevier Science, pp. 3297–3380.

Calvó-Armengolo, A., Patacchini, E., & Zenou, Y. (2009). Peer effects and social networks in education. *Review of Economic Studies*, *76*, 1239–1267.

Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, *81*(3), 855–882.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *Journal of Human Resources*, *45*(3), 655–681.

Dahl, B. G., Løken, K. V., & Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, *104*(7), 2049–2074.

De Giorgi, G., Fredriksen, A., & Pistaferri, L. (2016). Consumption network effects. (*NBER Working Paper No.22357*). Cambridge, MA: National Bureau of Economic Research.

De Giorgi, G., Pelizzari, M., & Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, *2*, 241–275.

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review Papers and Proceedings*, *95*(2), 158–165.

Dee, T. S. (2007). Teachers and the gender gaps in student achievement. *Journal of Human Resources*, *42*(3), 528–554.

Del Bello, C. L., Patacchini, E., & Zenou, Y. (2015). Neighborhood effects in education. (*IZA Discussion Paper No.8956*). Bonn, Germany: Institute of Labor Economics.

Del Boca, D., Monfardini, C., & Nicoletti, C. (2017). Parental and child time investments and the cognitive development of adolescents. *Journal of Labor Economics*, *35*(2), 565–608.

Dolton, P. (2017). Identifying social network effects. *Economic Record*, *93*(Supplement S1), 1–15.

Dustan, A. (2018). Family networks and school choice. *Journal of Development Economics*, *134*, 372–391.

Epple, D., & Romano, R. (2011). Peer effects in education: A survey of the theory and evidence, *Handbook of social economics* (Vol. 1, pp. 1053–1163) .Elsevier.

Ewin Smith, T. (1993). Growth in academic achievement and teaching younger siblings. *Social Psychology Quarterly*, *56*(1), 77–85.

Fletcher, J., Hair, N. L., & Wolfe, B. L. (2012). Am I My brother's keeper? Sibling spillover effects: The case of developmental disabilities and externalizing behavior. (*NBER Working Paper 18279*). Cambridge, MA: National Bureau of Economic Research.

Gibbons, S., Silva, O., & Weinhardt, F. (2013). Everybody needs good neighbours? Evidence from students' outcomes in England. *Economic Journal*, *123*(571), 831–874.

Gibbons, S., & Telhaj, S. (2016). Peer effects: Evidence from secondary school transition in England. *Oxford Bulletin of Economics and Statistics*, *78*(4), 548–575.

Hahn, J., & Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica*, *70*, 163–189.

Hobbs, G., & Vignoles, A. (2010). Is free school meal "eligibility" a good proxy for family income? *British Educational Research Journal*, *36*(4), 673–690.

Joensen, J. S., & Nielsen, H. S. (2017). Spillovers in educational choice. Available at SSRN: https://doi.org/ssrn.com/abstract=2548702 or https://doi.org/10.2139/ssrn.2548702

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics*, *17*, 99–121.

Lavy, V., Silva, O., & 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.

Lee, L. F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews*, *22*, 307–335.

Lindahl, L. (2011). A comparison of family and neighbourhood effects on grades, test scores, educational attainment and income: Evidence from Sweden. *Journal of Economic Inequality*, *9*(2), 207–226.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*, 531–542.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, *14*, 115–136.

Mazumder, B. (2008). Sibling similarities and economic inequality in the U.S. *Journal of Population Economics*, *21*(3), 685–701.

Moffitt, R. (2001). Policy interventions, low-level equilibria, and social interactions. In Durlauf, S. H., & Peyton Young, H. (Eds.), *Social dynamics* (pp. 45–82). Cambridge, MA: MIT Press.

Nicoletti, C., & Rabe, B. (2013). Inequality in pupils' test scores: How much do family, sibling type and neighbourhood matter? *Economica*, *80*(318), 197–218.

Nicoletti, C., & Rabe, B. (2018). The effect of school spending on student achievement: Addressing biases in value-added models. *Journal of the Royal Statistical Society, Series A*, *181*(2), 487–515.

Nicoletti, C., Salvanes, K., & Tominey, E. (2018). The family peer effect on mothers' labour supply. *American Economic Journal: Applied Economics*, *10*(3), 206–234.

Oettinger, G. S. (2000). Sibling similarity in high school graduation outcomes: Causal interdependency or unobserved heterogeneity? *Southern Economic Journal*, *66*(3), 631–648.

Patacchini, E., Rainone, E., & Zenou, Y. (2011). Dynamic aspects of teenage friendships and educational attainment. (*CEPR Discussion Paper No.DP8223*). London, UK: Centre for Economic Policy Research.

Qureshi, J. A. (2018a). Additional returns to investing in girls' education: Impact on younger sibling human capital. *Economic Journal*, *128*(616), 3285–3319.

Qureshi, J. A. (2018b). Siblings, teachers, and spillovers on academic achievement. *Journal of Human Resources*, *53*(1), 272–297.

Raaum, O., Salvanes, K. G., & Sørensen, E. (2006). The neighborhood is not what it used to be. *Economic Journal*, *116*(508), 200–222.

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. A., Machin, S., & Woessmann, L. (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 249–277). Amsterdam, Netherlands: Elsevier.

Slater, H., Davies, N. M., & Burgess, S. (2010). Do teachers matter? Measuring the variation in teacher effectiveness in England. *Oxford Bulletin of Economics and Statistics*, *74*(5), 629–645.

Todd, P., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, *113*(485), F3–F33.

How to cite this article: Nicoletti C, Rabe B. Sibling spillover effects in school achievement. *J Appl Econ*. 2019;1–20. https://doi.org/10.1002/jae.2674