

# Cognitive and Non-Cognitive Impacts of High-Ability Peers in Early Years\*

CHRISTOPHER BELFIELD† and IMRAN RASUL‡

†*Institute for Fiscal Studies*  
(christopher.belfield@ntlworld.com)

‡*Institute for Fiscal Studies; University College London*  
(i.rasul@ucl.ac.uk)

## Abstract

The sorting of students into ability groups is one of the most common, controversial and long-examined educational practices. Ability grouping also mechanically changes peer groups. We provide novel evidence on the cognitive and non-cognitive impacts in early years, of being exposed to higher-ability classroom peers through being assigned to the top within-class ability group. We exploit panel data from the UK Millennium Cohort Study, which allows us to construct trajectories of the cognitive and non-cognitive development of children from birth to entry into primary school. The data also record school grouping policies and the specific within-class group assignment of each child, by subject. We combine these rich data with an instrumental variable design using child-level variation in group assignment due to month of birth, in order to measure the local average treatment effect (LATE) of being assigned to the highest-ability peer group. We find that if a marginal student is assigned higher-ability peers, this significantly reduces their cognitive achievement in

\* Submitted January 2020.

We gratefully acknowledge financial support from the ESRC through the Centre for the Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (grant number ES/M010147/1). We thank the editor, two anonymous referees, Oriana Bandiera, Pedro Carneiro, Damon Clark, Bentley Macleod, Peter Fredriksson, Christopher Nielson, Kjell Salvanes, Michela Tincani, Miguel Urquiola, Daniel Wilhelm and numerous seminar participants for valuable comments. All errors remain our own.

Keywords: early years, peers, cognitive, non-cognitive outcomes.

JEL classification numbers: I21, I24.

mathematics, and has no impact on literacy. There are countervailing impacts on non-cognitive outcomes for the marginal student assigned higher-ability peers: although they are more motivated to study and parents respond with an improved home learning environment, these children have more study-related difficulties and their relations with their peers significantly worsen. These findings have important policy implications for the use and design of within-class ability grouping in early years.

## I. Introduction

Sorting school children into ability groups is one of the most common, controversial and long-studied educational practices.<sup>1</sup> Ability grouping can take many forms, ranging from within-class grouping where children are physically in the same classroom but separated into ability groups, to tracking systems where children are assigned to classes based on ability. Ability grouping also mechanically changes peer groups. How this affects inequality between groups depends on the nature of peer effects across the ability distribution.<sup>2,3</sup>

We study the effects of being exposed to higher-ability classroom peers through the use of within-class ability groups in UK primary schools. We exploit panel data on a nationally representative cohort of over 4,000 children from the Millennium Cohort Study (MCS) that has four key features. First, the MCS includes teacher surveys that record the grouping policy of the child's school, and the specific within-class group assignment of the child, by subject. School grouping (and tracking) policies are often informally adopted, and so there are loose definitions of such practices in most existing data, and information on the specific group assignment of a child is even rarer.<sup>4,5</sup> Second, cognitive tests of mathematics and literacy ability at age 7 are administered as part of the MCS surveys: these are taken at different times by children, allowing for age-at-test effects to be controlled for when studying these cognitive outcomes. Third, the MCS contains a rich array of non-cognitive outcomes related to teacher–child relations, home environment, socio-emotional development and peer relations. This allows us to build a holistic picture of the impacts on children of exposure to high-ability peers

<sup>1</sup>Oakes, 2005.

<sup>2</sup>Epple and Romano, 2011.

<sup>3</sup>With linear peer effects, grouping increases the variance in outcomes with ambiguous effects on achievement. With non-linear peer effects, high-ability students lose nothing, while low-ability students lose interaction with high-ability peers (Zimmer, 2003; Ding and Lehrer, 2007). If positive peer effects are strongest among similar types, grouping can improve the performance of all and potentially reduce inequality in achievement.

<sup>4</sup>Figlio and Page, 2002.

<sup>5</sup>As a result, typically researchers have had to resort to indirectly identifying sorting based on students' earlier test scores and comparing class assignments to perfect sorting or random assignment.

in early years, which is a critical time when cognitive and non-cognitive skills are being formed.<sup>6</sup> Fourth, it follows children from birth to early years schooling, allowing the construction of trajectories of cognitive and non-cognitive development of children from birth through to school entry (i.e. the point at which within-class group assignment decisions are made by teachers).

We combine these rich panel data with a research design based on instrumental variables (IV), exploiting variation in a child's ability-group assignment, driven by their month of birth, to identify the local average treatment effect (LATE) on the marginal child just assigned to the top-ability within-class group. Our research design measures the LATE of being assigned high-ability peers, for complier children whose top-group assignment is affected by their month of birth, conditional on their ability trajectory since birth, and all else equal.

Figure 1 shows the prevalence of ability grouping practices across countries using data from the OECD's Programme for International Student Assessment (PISA). Nearly every country uses some form of grouping, with 55 per cent of older pupils in OECD countries being exposed to within-class grouping. The prevalence is higher in the UK and US.<sup>7</sup>

Ability grouping generates heated debate, because of many perceived benefits and costs of such practices, the peer effects generated, and disagreement over whether the net effects aggravate economic inequality.<sup>8,9</sup> Given the prevalence of within-class grouping in UK primary schools, we do not exploit differences between schools with and without ability grouping. There are separate and well-established bodies of literature that measure the impact of ability grouping relative to no such grouping,<sup>10</sup> or that have studied the impacts of between-class ability tracking relative to no such tracking,<sup>11</sup> or

<sup>6</sup>Cunha, Heckman and Schennach, 2010.

<sup>7</sup>The prevalence of grouping in US schools is well noted using other data sources (Loveless, 2013; Dieterle et al., 2015). Figure 1 highlights that the other common form of grouping in later years is between-class tracking, where pupils are grouped by ability in a subject (so their peers might differ across subjects). This type of tracking is more common in larger, secondary schools, and it is not as relevant for early years.

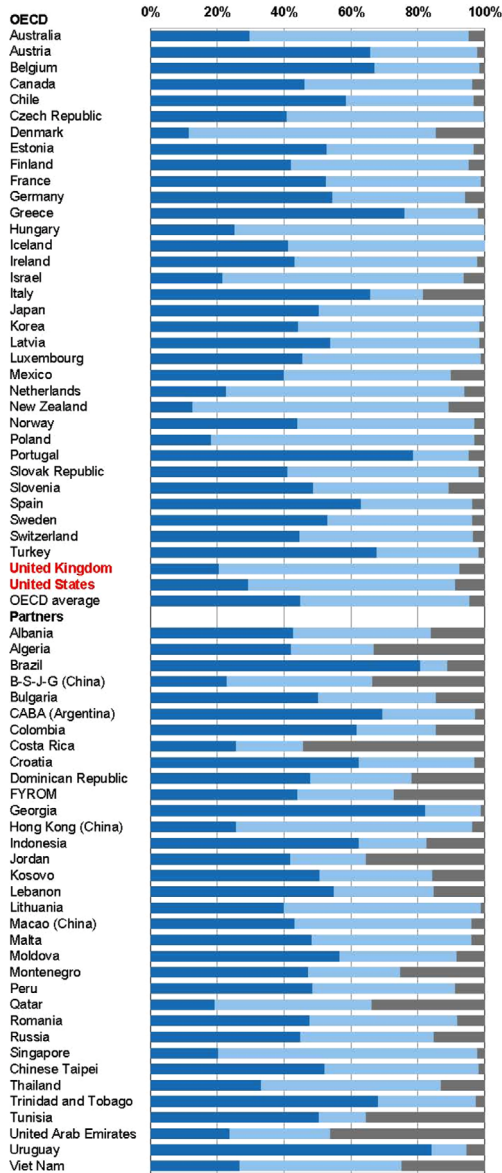
<sup>8</sup>Hanushek and Wößmann, 2006.

<sup>9</sup>Benefits include tailored pedagogy/resources to homogeneous ability groups, which in turn can create a virtuous cycle by avoiding student boredom and discouragement, and increasing children's motivation. Costs include lowered expectations and inputs towards lower-ability group students, leading to a vicious cycle where children's motivation is weakened and there is a self-fulfilling prophecy of low achievement. Using a large representative US sample of school resources, Betts and Shkolnik (2000a) and Rees, Brewer and Argys, (2000) find that for schools using between-class tracking, less experienced or less qualified teachers are assigned to lower tracks. This channel is less relevant for within-class grouping where students are taught in the same class, typically led by the same teacher.

<sup>10</sup>Slavin, 1990; Hoffer, 1992; Argys, Rees and Brewer, 1996; Betts and Shkolnik, 2000a, 2000b; Betts, 2011.

<sup>11</sup>Duflo, Dupas and Kremer, 2011; Vardardottir, 2013; Card and Giuliano, 2016.

FIGURE 1  
Ability grouping policies by country, PISA 2015 data



Note: The figure shows the percentage of students in schools where students are grouped by ability within classes: ■ no ability grouping for any subject; ■ one form of grouping for some subjects; ■ one form of grouping for all subjects.

Source: OECD, PISA 2015 Database, Tables II.5.14, II.5.22, II.5.27.

that have exploited tracking policy reforms to measure the impact of changes in a bundle of features of tracking systems on child outcomes.<sup>12,13</sup>

In contrast, our LATE measures the impact of a child being assigned high-ability peers in the top group, relative to outcomes if the child had been assigned to the next within-class ability group. As such, we bridge between the literature on ability grouping and on ability peer effects in classrooms. Earlier studies based on early years include Hoxby (2000), Hanushek et al. (2003), Angrist and Lang (2004), Lefgren (2004), Ammermueller and Pischke (2006) and Neidell and Waldfogel (2010).<sup>14</sup> Our analysis complements this evidence by using a nationally representative cohort of 4,000 children in primary schools in the UK, studying a rich set of cognitive and non-cognitive outcomes, and exploiting a traditional within-class grouping policy to identify the impact of high-ability peers on the marginal child assigned to top groups.

The key empirical challenge is to address the endogenous assignment of children to ability groups. Our research design uses the idea that a child's human capital can be decomposed into two time-varying dimensions: (i) productive skills that have a causal impact on attainment, denoted  $\theta_{it}$ ; (ii) non-productive traits that have no causal impact on later educational attainment ( $v_{it}$ ). Productive skills  $\theta_{it}$  can be serially correlated within a child over time. Non-productive traits  $v_{it}$  are more transient, but still salient to teachers in period  $t$ . To proxy teacher's information on past, current and projected productive

<sup>12</sup>Meghir and Palme, 2005; Hanushek and Wößmann, 2006; Pischke and Manning, 2006.

<sup>13</sup>The key challenge is schools' endogenous choice of grouping practices and assignment of children to groups. Early designs relied on variation between schools with and without a given policy, in conjunction with IV/matching methods, to deal with both layers of endogeneity (Hoffer, 1992; Argys et al., 1996; Betts and Shkolnik, 2000b). Much of this evidence is from secondary schools, so these designs are subject to further concerns, because in later years observing children's ability at school entry is insufficient to account for all the information teachers use in assigning children to groups (Betts and Shkolnik, 2000a; Pischke and Manning, 2006; Betts, 2011). A small set of older randomized controlled trials on ability grouping exists, although these have used small samples and have designs that often shut down mechanisms, or are known to be short term and so prevent equilibrium adjustments on other margins (Slavin, 1990). On tracking, Card and Giuliano (2016) use a research design based on the introduction of a gifted high-achiever programme in a large US school district. This required schools to introduce a top-ability class for fourth/fifth graders if they had at least one such gifted child in the school. These top groups are then usually filled by non-gifted high achievers. As the programme assigns students to the top class group based on test scores, it permits a regression discontinuity design, comparing students by ability rank at the margin between groups. They find that participation in the high-achiever group raises achievement for non-gifted high achievers among minorities, the key mechanism being improved teacher expectations and the removal of negative peer pressure for minority students (rather than teacher quality or ability-driven peer effects). Vardardottir (2013) exploits a somewhat similar fuzzy regression discontinuity design on a sample of teenagers using data from Iceland, to study the effects of assignment to high-ability classes. Finally, Duflo et al. (2011) implement a randomized controlled trial on tracking in elementary schools in Kenya.

<sup>14</sup>Ability peer effects in other parts of the education system well after early years have been studied. Boojj, Leuven and Oosterbeek (2017) present experimental evidence from university students, and find that homogeneous ability groups lead to better test-score outcomes. Garlick (2018) presents evidence from students in a South African university, exploiting students who are either tracked or randomly assigned to dorms in order to understand peer effects on grades.

skills (i.e. the entire vector of productive skills over time  $\theta_i$ ), we use the MCS panel structure to construct trajectories of the pre-school development for a child based on their birthweight, cognitive and non-cognitive ability as measured at ages 1, 3 and 5. These ability trajectories allow us to condition on: (i) a richer set of cognitive ability measures than those contemporaneously available at the point of school entry; (ii) trajectories of non-cognitive skills of the child, encompassing personality traits and other behaviours.

Our research design assumes that teachers observe the wealth of information that goes into these dimensions of child development since birth and ages 1, 3 and 5, when making decisions about group assignment. These dimensions correspond to  $\theta_i$ , the observed and projected ability of child  $i$ , and are assumed to encompass all true measures of a child's developmental readiness. Conditional on this ability trajectory from birth to school entry, and other child, family and school controls, our IV strategy then isolates variation in non-productive traits of children that teachers treat as signals of ability, when children are initially assigned to within-class groups. The instrument used to generate such variation is a child's month of birth. An example of a non-productive and transient trait we have in mind is how boisterous a child is at the time when decisions about group assignment are made. Teachers can be attentive to such non-productive traits when group-assignment decisions are made, even though such transient traits are not predictive of the child's true ability or later attainment.

We assume that initial group-assignment decisions are sticky, so that a child's true ability is only slowly revealed. Hence, children are exposed to the ability group they are initially assigned to, for some time thereafter. Indeed, a concern often raised about within-class grouping is the lack of mobility across groups, which leads to a misallocation of students to ability groups. This concern is reinforced if initial group assignment is based on noisy proxies of ability.<sup>15</sup>

The validity of the IV depends on the exclusion restriction. This requires that conditional on a child's ability trajectory from birth to school entry, noisy signals of ability embodied in month of birth, which only affect attainment at age 7 through group assignment, and exposure to classroom peers of different ability. Of course, an established literature has documented month-of-birth effects on attainment.<sup>16</sup> The potential channels proposed for this are: (i) age-at-test; (ii) age at which a child starts school (school readiness); (iii) the length of schooling (or exposure to the home environment); (iv) relative age effects, including differential attention from teachers.

<sup>15</sup>Evidence on this for the UK is limited: Barker Lunn (1970) suggests that 15 per cent of children end up in the wrong track by the end of the school year, based on English and arithmetic performance.

<sup>16</sup>Crawford, Dearden and Greaves, 2014.

Separating out these factors is challenging because children typically sit school exams at the same time of year, so factors (i)–(iii) above are mechanically linked. This is not the case in the MCS data: we measure cognitive attainment in mathematics and literacy at age 7 using tests administered as part of the MCS, so taken by children at different ages. Hence, we condition on age-at-test directly in our first and second stages. Moreover, our ability trajectory measure  $\theta_i$  includes controls for school readiness at age 5.

Once these month-of-birth related factors (age-at-test, school readiness at age 5, ability trajectories from birth) are conditioned out, this only leaves month of birth as potentially capturing a relative age effect, which can violate the exclusion restriction. We address this issue by showing the impact of being assigned more able peers in top groups on outcomes such as teacher–child interactions, child motivation, socio-emotional development and relationship with classroom peers. All of these are claimed to be mechanisms for relative age effects, and are generally beneficial for children who are older than their classroom peers.<sup>17</sup>

The UK school year starts on 1 September and children are eligible to start primary school if they turn 5 in the relevant school year. Hence, the oldest children in a cohort are those born in September, and the youngest are those born in August. As we show later, delaying entry into school (red-shirting) occurs infrequently in the UK, as confirmed in the MCS data.

Our main results on the impact on the marginal child assigned to top-group ability peers are as follows. In the first stage, month of birth is highly predictive of group assignment in both subjects. In mathematics, relative to those born in the first term (September to December), those born in the second term are 9 percentage points (pp) less likely to be assigned to the top group ( $p < 0.001$ ); those born in the final term (May to August) are 18 pp less likely to be assigned to the top group ( $p < 0.001$ ). The  $F$ -statistic is over 20, showing the strength of the instrument conditional on ability trajectories from birth to age 5, age-at-test, child, household and school characteristics. The conditional likelihood to be assigned to the top group is significantly different for all three terms of birth, so the instrument shifts children from multiple parts of the term of birth distribution into top-group assignment. A very similar set of first stage results emerge for assignment to the top group in literacy.<sup>18</sup>

In the second stage, once endogenous assignment is accounted for, the impacts on attainment (measured in effect sizes) are either negative for mathematics ( $\hat{\beta}_{IV} = -0.136$  sd,  $p < 0.001$ ) or not statistically different from

<sup>17</sup>Bedard and Dhuey, 2006.

<sup>18</sup>There has been a long debate on the validity of IV designs for the causal impact of educational attainment on later life outcomes, using month/season of birth as an instrument; see Bound, Jaeger and Baker (1995). Much of this has emphasized the problem of weak instruments. In contrast, we use month of birth as an instrument for group assignment (not attainment) and, as documented below, our first stage suffers less from weak instrument concerns.



zero for literacy ( $\hat{\beta}_{IV} = -0.013$  sd). Thus, the OLS and IV estimates are of opposite sign. As expected, the OLS estimate is upward biased because group assignment captures elements of a child's ability,  $\theta_i$ , that drive later attainment. The IV estimate is negative: the interpretation is that the marginal child assigned to the top group, because the teacher uses non-productive traits correlated to month of birth ( $v_{it}$ ) as signals of ability at the time of initial group assignment, does significantly worse in mathematics attainment at age 7, relative to the counterfactual of having been assigned outside the top group. The 95 per cent confidence interval on the IV estimate rules out a causal effect size of top-group assignment for mathematics (literacy) larger than  $-0.083$  sd (0.031 sd).

This negative impact occurs despite the fact that these marginal children are older (to reiterate from the first stage, older children are significantly more likely to be assigned to the top group, all else equal), and so these children are of higher relative age as well. This casts doubt on the concern that month of birth only captures relative age impacts (because that should lead to marginal children having higher cognitive achievement, all else equal).

On non-cognitive outcomes, we identify countervailing costs and benefits to the marginal child of being quasi-randomly assigned to the top within-class ability group. On the one hand, this causes children to be more motivated to study, and parents respond with an improved home learning environment. On the other hand, teachers to report the child as having more study-related difficulties, and relations with peers to worsen on multiple margins (children are more likely to be solitary, and to be bullied and fight with others).

This array of offsetting non-cognitive impacts suggests why ability grouping generates so much controversy: the strength of mechanisms likely varies across children, households and schools. This inevitably leads to diverging opinions as to the educational value of such policies.<sup>19</sup>

Our contributions are to study the impacts of higher-ability peers in the context of within-class ability grouping in early years, combining a research design and rich panel data that allow us to study these impacts on marginal children assigned to top groups based on variation in their month of birth. Grouping practices, unlike other school-based inputs, can be changed using mostly existing resources. There are thus high returns to making correct decisions over ability grouping. This is especially so in early years, a critical

<sup>19</sup>Our results have implications for regression discontinuity designs studying the impact of school quality on child outcomes, where marginally admitted and rejected students are compared; see Cullen, Jacob and Levitt (2006), Hoekstra 2009, Jackson 2010, Dobbie and Fryer 2011, Abdulkadiroglu, Angrist and Pathak (2014), Clark and Del Bono (2016). There might well be a rich set of effects across cognitive and non-cognitive dimensions in such settings also.



time for the production of human capital, with skills accumulated being complementary to later learning.<sup>20,21</sup>

In Sections II and III, we describe the data and research design. In Sections IV and V, we present the main results on cognitive and non-cognitive outcomes. In Section VI, we discuss further policy implications.

## II. Data

The MCS is a panel data set following a cohort of 19,000 children born in the UK in 2000/01. Information has been collected on this cohort at birth, and every two/three years subsequently. We focus on the 13,847 children followed to age 7, when they are all in primary school.<sup>22</sup> In the age 7 survey wave, cohort members' teachers are interviewed. This enables us to measure the ability-grouping policies of the school (specifically those applicable to the child's school year), and the specific group assignment of the child. Obtaining this information directly from teachers is important because parents and pupils often have ambiguous views on what constitutes ability grouping.<sup>23,24</sup>

The teacher survey also provides information about teacher-child interactions, the child's behaviour and attitudes towards school. There are 8,765 responses to the teacher survey. Columns 1 and 2 of Table A.1 in the online appendix compare characteristics of children, households and schools between those in the baseline MCS sample, and those with a teacher survey. The samples can be compared well on observables, although, to reiterate, our research design never exploits between-school differences.<sup>25</sup>

<sup>20</sup>Cunha et al., 2010.

<sup>21</sup>Early years are not too soon for the effects to be detectable: studies have shown sizeable effects of school inputs on attainment in early years, such as for class size (Krueger, 1999) and teacher quality (Chetty, Friedman and Rockoff, 2014).

<sup>22</sup>We narrow the sample to children born in the 2000/01 academic year. All children in the MCS from England and Wales were born between 1 September 2000 and 31 August 2001, so this restriction does not apply to them; see Hallam and Parsons (2013). Children from Scotland and Northern Ireland were born between 1 November 2000 and 14 January 2002.

<sup>23</sup>Rosenbaum, 1980; Rees et al., 2000.

<sup>24</sup>The wording of the MCS question defining ability grouping allows teachers to report multiple types of policy (including within- and between-class grouping) and whether they apply to specific or to all subjects (streaming); see Parsons and Hallam (2014). The wording of the opening question is as follows. 'We are interested to know about groupings between and within classes in this child's year. Some schools group children in the same year by general ability and they are taught in these groups for most or all lessons. We refer to this as streaming. Some schools group children from different classes by ability for certain subjects only and they may be taught in different ability groups for different subjects. We refer to this as setting. Other schools do not group children by ability between classes. Sometimes this may be because there are not multiple classes in the year.'

<sup>25</sup>In both samples, around 85 per cent of pupils are white and there is an equal split across genders. Those in the sample with a teacher survey are from households with slightly higher income and are less likely to reside in England.

TABLE 1  
*Within-class ability grouping policies*

	All schools (1)	Also use between-class setting (2)	Do not use between-class setting (3)
Percentage of schools using within-class ability groups	85%	27%	58%
Percentage of children in within-class ability groups	86%	28%	57%
<i>Means (std dev. in parentheses)</i>			
Among schools using within-class ability grouping:			
Mathematics only	21%	23%	11%
Literacy only	9%	9%	6%
Both mathematics and literacy	71%	68%	83%
	100%	100%	100%
Number of within-class ability groups:			
Mathematics	3.59 (0.928)	3.57 (0.969)	3.60 (0.910)
Literacy	3.77 (0.923)	3.76 (0.953)	3.77 (0.908)

*Note:* This shows the prevalence of within-class ability groups and between-class setting as recorded in the MCS teacher survey. Within-class ability grouping refers to the sorting of students into different working groups, based on their ability, within the same physical class. Between-class setting is defined as sorting students, based on their ability, into different classes.

The MCS allows us to study the impacts of high-ability peers using a sample of children drawn from 4,000 primary schools nationwide, moving beyond the context-specific policies of particular schools or jurisdictions. The cost is that the data are not well suited to examine peer effects directly, as we only observe a few children per school. The strength of the data is that we can examine the impact of high-ability peers on a range of non-cognitive outcomes, such as teacher–child interactions, child development and peer relations. All these could mediate the impact of high-ability peers on cognitive outcomes.

### 1. Measuring ability grouping

Table 1 shows the use of ability grouping in early years, as reported by teachers in the MCS teacher survey. Column 1 shows that 85 per cent of schools use within-class grouping (with 86 per cent of children being so grouped). Column 2 shows that just over a quarter of these schools also use between-class tracking (where children are further taught by ability groups in separate

classes), but column 3 shows that the majority of primary schools exclusively use within-class grouping. We focus our analysis on schools using only within-class grouping because this allows for a cleaner identification of the impacts of exposure to high-ability peers, without being confounded by other ability grouping policies.<sup>26</sup>

The lower panel of Table 1 reveals subject-specific information. Most schools use within-class grouping for mathematics and literacy. There are around four within-class groups for each subject. Within-class grouping typically involves physically placing children on to separate tables in the same classroom, with teachers assigning different work to children on each table. Teachers, and teaching assistants, can move across tables, providing inputs to all groups. The group assignment of a child can vary by subject: just under a quarter of children are in different ability groups for mathematics and literacy. Given the relatively small size of most primary schools, usually the same teacher will teach both subjects.

With average class sizes in primary school of 26 in 2007 and four groups per subject, around seven children are assigned to each group. One of these will (implicitly) be designated a top group with the highest-ability children. We could evaluate the impact of being marginally assigned to any given group relative to others. However, given variation across schools in the number of groups per subject, we focus on a comparison that can be made in all schools: that is, between the assignment of the marginal child to the top ability group (with the most able peers) relative to their assignment to the next ability group (with lower-ability peers on average).

## 2. Measuring ability trajectories

The panel dimension of the MCS allows us to construct the ability trajectory of each child from birth to school entry. We use birthweight to proxy initial endowment at birth. We construct index measures of the child's cognitive ability at ages 1, 3 and 5, and the child's non-cognitive ability/socio-emotional development at ages 1, 3 and 5.<sup>27</sup>

<sup>26</sup>Columns 3 and 4 of Table A.1 show how child, household and school characteristics differ across primary schools that do and do not use within-class grouping. We see few large differences on observables, although schools using within-class grouping are more likely to be in England, have larger class sizes, and are less likely to have mixed year classes. Column 5 shows descriptives for those schools exclusively using within-class grouping (so not using between-class tracking). A comparison of columns 3 and 5 again reveals few notable observable differences across these kinds of school. To reiterate, our research design never exploits these between-school differences.

<sup>27</sup>An established literature documents that better neonatal health, proxied by birthweight, has positive impacts on long-run cognitive outcomes, such as educational attainment, IQ and earnings; see Black, Devereux and Salvanes (2007), Figlio et al. (2014) and Dhuey et al. (2017). Such channels are thus conditioned out in our design. We later show the robustness of our main finding to controlling for additional physical traits of children.

On the cognitive ability measures, at age 1 the index is constructed from factors reported by parents. Cognitive ability measures at age 3 are MCS enumerator assessments of the child, where again multiple factors go into the assessment. Cognitive ability at age 5 is measured using the British Ability Scale (BAS) tests on vocabulary, picture and pattern recognition, where each component is equally weighted. On the non-cognitive ability measures, at age 1 the index is constructed from behavioural factors of the child reported by their parents. At ages 3 and 5, the non-cognitive development of the child is measured through a socio-emotional index of development, based on parental reports.<sup>28</sup>

Our research design assumes that teachers observe these dimensions of child development since birth, ages 1, 3 and 5 when making decisions about group assignment. These dimensions correspond to  $\theta_i$ , the observed and projected ability of child  $i$ , and are assumed to encompass all true measures of a child's developmental readiness. Conditional on these, we assume that teachers, on the margin, base group-assignment decisions on a child's month of birth because this is correlated to non-productive and transient traits ( $v_{it}$ ) that they treat as signals of ability.

Cognitive outcomes are measured at age 7. These are also derived from cognitive tests administered as part of the MCS age 7 survey wave. The first test measures mathematics achievement, and is a test devised by the National Foundation for Educational Research (NFER). The second measures literacy achievement and is the BAS test on reading and pattern recognition.

<sup>28</sup>The factors used to construct the ability index at age 1 are the following: whether the child sits up, stands up while holding on, moves their hands together, grabs objects, holds small toy, can walk, can give a toy, can wave, can extend their arms, can nod. The age 3 cognitive assessment is based on MCS enumerator administered tests on colours, letters, numbers, sizes, comparisons, shapes, and naming vocabulary. Confirmatory factor analysis is used to construct the ability indices at age 1 and age 3 (where the appropriate weights are calculated in the factor analysis according to the amount of information in the measure). On the non-cognitive ability measures, the index at age 1 is constructed from a battery of parental reported behaviours of the child, including the child makes happy sounds (coos, laughs), is pleasant (smiles, laughs) when first arriving in an unfamiliar place, is pleasant during procedures such as hair brushing or face washing, is content during interruptions of milk or solid feeding, remains pleasant or calm with minor injuries, objects to being bathed in a different place or by a different person, still wary or frightened of strangers after 15 minutes, shy on meeting another child for the first time, frets for the first few minutes in a new place or situation, appears bothered when first put down in a different sleeping place, wants to take milk feeds at about the same time every day, gets sleepy at the same time each day, naps about the same time each evening, naps about the same length day to day, and wants to take solid foods at about the same time day to day. The non-cognitive/socio-emotional index at age 3 is constructed from a battery of behavioural questions reported by parents, which cover the following domains: independence, emotional dysregulation, prosociality, conduct, hyperactivity, and emotional skills. The same domains are covered at ages 5 and 7 but also include questions on cooperation. Confirmatory factor analysis is used to construct the non-cognitive/socio-emotional indices at ages 1, 3, 5 and 7 (where weights are calculated in the factor analysis according to the amount of information in the measure).

We standardize mathematics and literacy test scores (based on the entire MCS sample), so coefficient effects represent standard deviation effect sizes.<sup>29</sup>

Month of birth is recorded for all children (exact date of birth is unavailable). The age at which each child takes the MCS-administered ability tests at age 5 and age 7 is also recorded. As children in each survey wave are interviewed on different dates, this allows us to control for age-at-test effects, both in terms of the main outcomes at age 7 and in terms of constructing an age-adjusted ability trajectory to school entry.

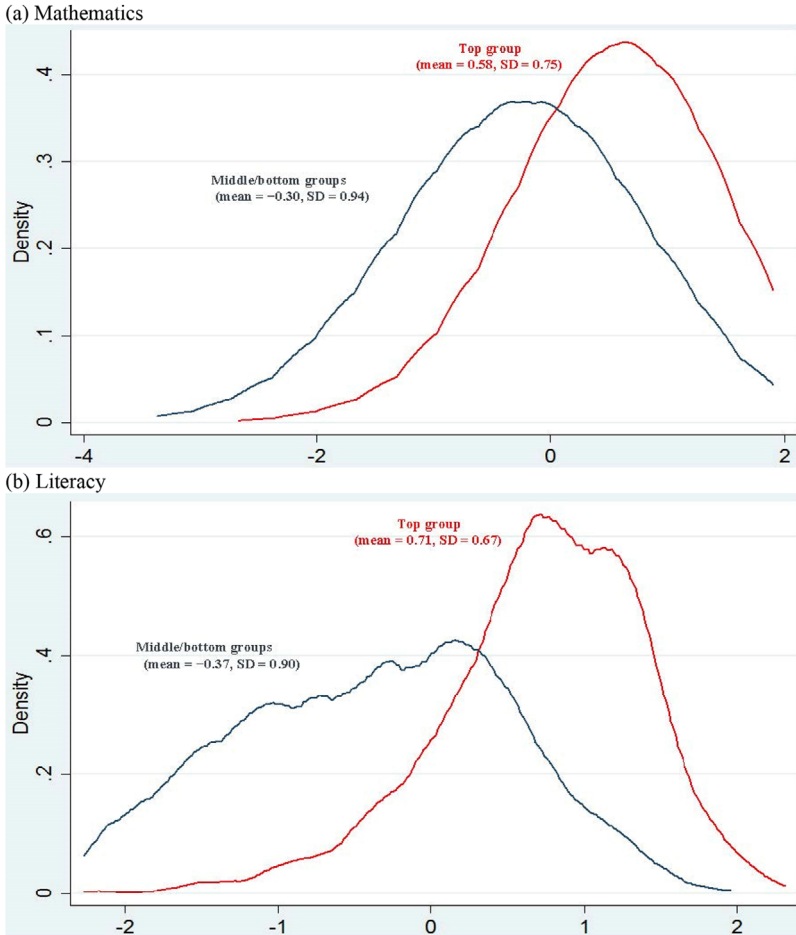
### 3. Descriptives

Panel A of Table A.2 in the online appendix presents descriptive evidence on peers by ability group. For both subjects, peer composition differs by group. In mathematics, top-group peers are more likely to be white or from higher-income households. In literacy, top-group peers are more likely to be girls or from higher-income households. On other dimensions of classroom characteristics, such as teachers and pedagogy, as shown in Panels B and C, almost by definition these are mostly common inputs to all within-class groups. For example, teachers and teaching assistants are common (and we later show that top-group children are no more likely to receive help from teachers or teaching assistants once their endogenous group assignment is accounted for). There is a statistically significant but small difference in the expected minutes of homework across groups (if anything, this should increase attainment in the top group).

Figure 2 show distributions of normalized test scores at age 7, by ability group and subject. Panel A reveals that, on average, children in the top within-class ability group for mathematics have a 0.88 sd higher test score than those in other groups, while in literacy the test score gap between the top group and other groups is 1.08 sd. The LATE measured by our research design is the test score impact for complier children whose assignment to the top group is affected by their month of birth, conditional on their ability trajectory.

<sup>29</sup>The BAS is an established procedure to assess a child's cognitive ability and educational achievement at early ages. The battery of individually administered tests is designed for use by educational and clinical psychologists to assess a child's ability, including picking up learning and behavioural difficulties. It provides a comprehensive means of assessing different aspects of a child's intellectual functioning from age 3 onwards. Nationwide administrative tests are conducted on children at age 7 in the UK. In our sample, this formal test score has a correlation of 0.43 (0.63) with the MCS-based attainment measure in mathematics (literacy).

FIGURE 2  
Attainment at age 7, by ability group



Note: The figures show, by mathematics and literacy, the distribution of normalized attainment measures at age 7 for those in the top within-class ability group, and those assigned to the middle/bottom groups. Attainment is measured by the NFER maths test for mathematics and the BAS reading test for literacy.

### III. Research design

Our estimating equation for child  $i$  in subject  $j$  (mathematics, literacy) is

$$(1) \quad y_{ij}^7 = \beta_0 D_{ij}^7 + \sum_{k=0,1,3,5} \beta_{1k} P_i^k + \beta_2 A_i^7 + \beta_3 X_i + u_{ij}^7.$$

Here,  $y_{ij}^7$  is the cognitive outcome in subject  $j$  at age 7, and  $D_{ij}^7 = 1$  if child  $i$  is assigned to the top within-class group for subject  $j$ . Child  $i$ 's ability trajectory,  $\theta_i$ , is captured by  $P_i^k$ , comprising eight dimensions: birthweight, cognitive and non-cognitive ability at ages  $k$  ( $k = 5, 3, 1$ ), and age-at-test for the age 5 cognitive ability measure. Two points should be noted: (i) the correlation between components is low (it is only above 0.5 for two out of the 28 pairwise correlations); (ii)  $P_i^k$  captures the fact that the ability trajectory for some children is rising, whereas it is falling for others. We show the robustness of our results to alternative specifications for ability trajectories.

$A_i^7$  is the age-at-test for the age 7 attainment tests. As Table A.1 shows, because each survey wave is collected through the year, there is variation in age-at-test, and this is an important driver of cognitive test scores that we condition out.  $X_i$  is a vector of background child, family and school characteristics. The school characteristics controlled for relate to class size, year cohort size and mixed year groups. We do not condition further on class characteristics as these might be endogenous to the use of within-class ability grouping.  $u_{ij}^7$  is an error term and robust standard errors are calculated.<sup>30</sup>

The coefficient of interest is  $\beta_0$ , which is the effect of being assigned to higher-ability peers in the top group. We account for the endogeneity of group assignment by modelling this as

$$(2) \quad D_{ij}^7 = F(P_i, A_i^7, X_i, M_i),$$

where this is estimated using a probit model. Here,  $P_i$  is the eight-dimensional vector capturing the child's ability trajectory through ages  $k$  described above,  $A_i^7$  and  $X_i$  are also as defined above, and  $M_i$  is our instrument (i.e. child  $i$ 's month of birth, or their term of birth). The UK school year starts on 1 September and children are eligible to start primary school if they turn 5 in the relevant school year. Hence, when using term of birth we partition  $M_i$  into those born September to December, January to April, and May to August.

We simultaneously estimate the system of equations in (1) and (2) using limited information maximum likelihood (LIML), which allows for a probit functional form in the first-stage.

Month/term of birth is an instrument that predicts group assignment, and that is excluded from the second stage. The IV strategy assumes variation in month/term of birth  $M_i$ , parsing out  $P_i$ ,  $A_i^7$  and  $X_i$ , and it isolates variation

<sup>30</sup>The child and family characteristics captured in  $X_i$  are binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in household, the log of family net income, and the number of siblings.



in group assignment arising from teachers using non-productive and transient traits as signals of the child's true ability when group assignment decisions are made. We take as given that such initial group assignment decisions are sticky, so that children are exposed to the group they are initially assigned to. This is because a child's true ability is only slowly revealed, and/or teachers might be reluctant to overturn their initial assignment decisions. To the extent there is mobility across groups, or groups are assigned close to age 7, the impact of top-group assignment will be attenuated.

The fact that being born earlier in the school year is beneficial for attainment is well established.<sup>31</sup> This is unsurprising: in an education system in which children must be age 5 to enter, age differences are 20 per cent between the oldest and youngest children in the class. Two explanations for month of birth effects are: (i) age-at-test/school readiness; (ii) relative age effects.

As Crawford et al. (2014) describe, explanation (i) comprises an age-at-test effect, an age-of-starting school effect (those born just before the academic year cut-off may be disadvantaged by starting school younger than their peers and so are less school ready), and a length-of-schooling effect (those born just before the cut-off are exposed to school for less time/exposed to their home environment for longer). Separating out these factors is typically challenging because children often sit school exams at the same time of year and so these three factors are mechanically related.<sup>32</sup>

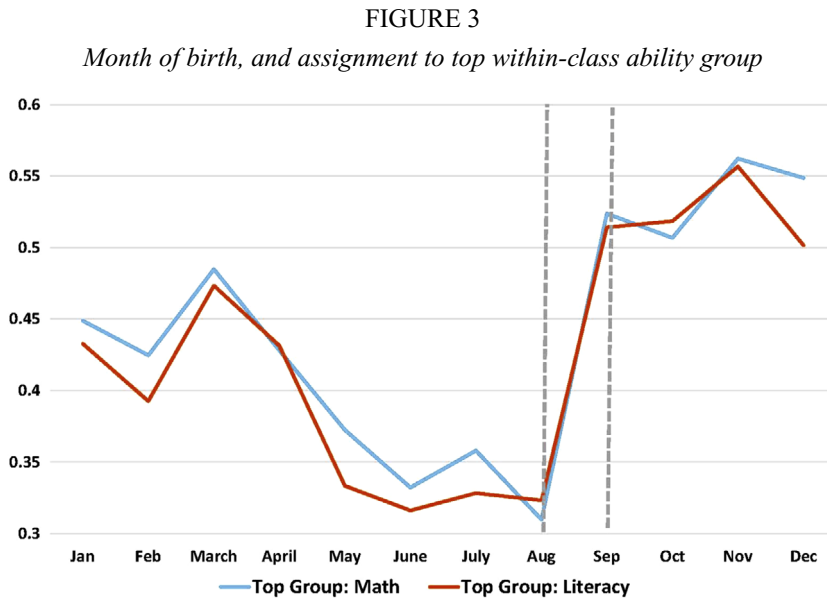
In our context, this is not the case: attainment at age 7 is measured in MCS-administered tests and so children vary in their age-at-test, and we can condition on this. Moreover, being able to follow children since birth allows us to condition on the school readiness of the child (as measured by their cognitive and non-cognitive development at age 5, as well as their age-at-test for these age 5 tests). This only leaves month of birth as potentially capturing a relative age effect. We later consider whether the evidence is consistent with a relative age effect (that tends to say older children within a class have better outcomes) for margins of impact, such as teacher-child interactions, child motivation, socio-emotional development and relationship with classroom peers.

Finally, we note that delaying entry into school (red-shirting) occurs infrequently in the UK context in general, and is measured to be so in the MCS.<sup>33</sup>

<sup>31</sup>Bedard and Dhuey, 2006; Black, Devereux and Salvanes, 2011; Fredriksson and Ockert, 2014.

<sup>32</sup>Crawford et al. (2014) find the vast majority of month-of-birth effects on test scores are driven by age-at-test. The combined effect of age of starting school, length of schooling and relative age is close to zero.

<sup>33</sup>In the MCS sample, less than 2 per cent of children have been held back, and as expected these are concentrated among those born in July and August (but still represent only 10 per cent of those born in those months). In the US, red-shirting is more common and would invalidate our research design (Dhuey et al., 2017).



*Note:* The figure shows the unconditional probability of being placed in the top or middle-bottom within-class ability group, by month of birth. The academic school year starts on 1 September.

## IV. Cognitive outcomes

### 1. First stage

Figure 3 shows how the unconditional likelihood of being assigned to the top within-class ability group varies by month of birth. The UK school year starts on 1 September, so the oldest children in a class are born in September. We observe a clear spike in assignment to the top ability group ( $D_{ij}^7 = 1$ ) for those born in September relative to those born in August. This spike occurs for assignment in mathematics and literacy. The majority of children born in the first term of birth (September to December) are assigned to the top ability group for mathematics and literacy. This falls to around one-third for those born in the last term of birth (i.e. May to August).

Table 2 shows the first-stage probit estimates by subject, reporting marginal effects. Focusing first on mathematics, Panel A (column 1) shows unconditional changes in probability of assignment to the top ability group across terms of birth. Relative to those born in the first term of birth, those born in the second term of birth are 9.5 pp less likely to be assigned to the top ability group ( $p < 0.001$ ); those born in the final term are 19 pp less likely to be assigned to the top ability group ( $p < 0.001$ ). Column 2 shows that these marginal effects remain stable once we condition on a rich set of covariates ( $\mathbf{P}_i, A_i^7, X_i^7$ ) and estimate equation 2 in full.

TABLE 2  
*Within-class ability grouping, first stage*

	Unconditional (1)	Baseline (2)	Excluding schools with between-class setting (3)
<i>Panel A: mathematics</i>			
Born January–April	−0.095*** (0.017)	−0.088*** (0.026)	−0.079* (0.032)
Born May–August	−0.190*** (0.017)	−0.180*** (0.035)	−0.170*** (0.042)
<i>F</i> -statistic	123	24.8	15.4
<i>p</i> -value (Jan–Apr = May–Aug)	0.000	0.000	0.000
Ability trajectory	No	Yes	Yes
Age-at-test	No	Yes	Yes
Child characteristics	No	Yes	Yes
Family background	No	Yes	Yes
School characteristics	No	Yes	Yes
Number of observations	5002	4341	3004
<i>Panel B: literacy</i>			
Born January–April	−0.095*** (0.016)	−0.106*** (0.026)	−0.108*** (0.030)
Born May–August	−0.188*** (0.016)	−0.209*** (0.033)	−0.237*** (0.038)
<i>F</i> -statistic	125	36.5	34.8
<i>p</i> -value (Jan–Apr = May–Aug)	0.000	0.000	0.000
Ability trajectory	Yes	Yes	Yes
Age-at-test	No	Yes	Yes
Child characteristics	No	Yes	Yes
Family background	No	Yes	Yes
School characteristics	No	Yes	Yes
Number of observations	5169	4482	3273

*Note:* \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . The dependent variable is assignment to the top ability group in a subject. Probit regressions and marginal effects. Robust standard errors are given in parentheses. The outcome variable is a dummy equal to one if the child is assigned to the top within-class ability group in mathematics (Panel A) and literacy (Panel B). We report marginal effects from a probity, of being born in the Spring term (January to April) and Summer term (May to August) on the probability of being assigned to the top within-class ability group. The omitted category consists of those children born in the Autumn term of the school year (September to December). Column 1 reports unconditional effects. Columns 2 and 3 condition on the following controls: the child's ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in the household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ 's birthweight, cognitive and non-cognitive ability at ages 1, 3 and 5, and their age-at-test for the age 5 cognitive ability measure. Column 3 excludes children in schools that use between-class setting. For each column, we report the *F*-statistic on the joint significance of the instrument (terms of birth) and the *p*-value on the null that the marginal effects for the two reported months of birth on assignment to the top group are equal.

For top-group assignment in mathematics, the  $F$ -statistic is over 24, showing the strength of the instrument even conditional on ability trajectories from birth to school entry, age-at-test, and child, household and school characteristics. The conditional likelihood to be assigned to the top group is significantly different for all three terms of birth, suggesting that the instrument affects treatment assignment in a way consistent with the monotonicity condition, and also shifts pupils from multiple parts of the term-of-birth distribution into the top within-class ability group for mathematics. Column 3 shows these first-stage findings to be robust in the sample of primary schools that only use within-class grouping (and so do not use between-class setting).

Panel B replicates the specifications for literacy: term of birth is even more strongly predictive of top ability group assignment for this subject ( $F > 34$ ), the conditional likelihood differs across the three terms of birth (with all  $p$ -values less than 0.001), and the first results are similar in schools with and without between-class setting.<sup>34</sup>

The evidence suggests that teacher's assignment algorithms when placing children into top within-class ability groups are similar across subjects, and between schools that do/do not additionally use between-class ability setting.<sup>35</sup>

## 2. Second stage

Table 3 shows the second-stage results, by subject. Panel A shows OLS estimates of equation 1. Panel B shows second-stage IV effects accounting for endogenous group assignment.

Column 1 shows the impact on mathematics attainment in the sample of all primary schools. The OLS estimate shows a strong positive partial correlation of test scores with top-group assignment:  $\hat{\beta}_{0,OLS} = 0.451$  sd ( $p < 0.001$ ). Accounting for endogenous group assignment, the IV estimate is  $\hat{\beta}_{0,IV} = -0.136$  ( $p < 0.001$ ). Hence, the OLS and IV estimates are of opposite sign. The IV estimate suggests that children who are quasi-randomly assigned to high-ability peers in the top group because teachers use traits correlated to month of birth as signals of ability, do significantly worse in mathematics at age 7, relative to the counterfactual of having remained with lower-ability peers outside the top group. This is despite the fact that these marginal children are

<sup>34</sup>Many of the components of  $P_i$  robustly predict top-group assignment across subjects and specifications (including birthweight, cognitive ability at ages 1, 3 and 5, and socio-emotional development at age 5).

<sup>35</sup>This result is useful because the instrument is invalid if month of birth is correlated to parental behaviour, such as lobbying to have their child placed into the top group. This might be picked up by the instrument having varying effects across schools with and without between-class setting. This is because between-class setting is easier for parents to observe and likely to have more effect on attainment than the use of within-class ability grouping alone.

TABLE 3  
*Within-class ability grouping and attainment, second stage*

	<i>Mathematics</i>		<i>Literacy</i>	
	All schools (1)	Excl. schools with between- class setting (2)	All schools (3)	Excl. schools with between- class setting (4)
<i>Panel A</i>				
Assigned to top group in subject	0.451*** (0.026) [0.400, 0.502]	0.445*** (0.032) [0.382, 0.507]	0.799*** (0.024) [0.752, 0.846]	0.806*** (0.029) [0.748, 0.864]
<i>Panel B</i>				
Assigned to top group in subject	-0.136*** (0.027) [-0.189, -0.083]	-0.118*** (0.030) [-0.176, -0.060]	-0.013 (0.022) [-0.056, 0.031]	-0.010 (0.025) [-0.059, 0.039]
Ability trajectory	Yes	Yes	Yes	Yes
Age-at-test	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes
Number of obs.	4341	3004	4482	3273

*Note:* \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . The dependent variable is attainment at age 7. OLS and LIML regressions. Robust standard errors are given in parentheses, and 95 per cent confidence intervals in brackets. Values show the effect of assignment to the top within-class ability group on attainment at age 7. Attainment is measured by the NFER math test for mathematics and the BAS reading test for literacy. The test scores are normalized so the results represent standard deviation effect sizes. Panel A presents OLS results for mathematics and literacy. Panel B reports limited information maximum likelihood results estimated using a conditional mixed process with term of birth omitted in the second stage. Attainment is measured by the NFER math test for mathematics and the BAS reading test for literacy. The test scores are normalized so the results represent standard deviation effect sizes. We condition on the following controls throughout: the child's ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in the household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ 's birthweight, cognitive and non-cognitive ability at ages 1, 3 and 5, and their age-at-test for the age 5 cognitive ability measure. Columns 2 and 4 exclude children in schools that use between-class setting.

older (and so of higher relative age) as the instrument is positively correlated with top-group assignment. For each estimate, we also report the 95 per cent confidence interval: the IV estimate rules out a causal effect size of top-group assignment on mathematics attainment any larger than  $-0.083$  sd.

Column 2 shows that all these findings continue to hold for children in the smaller sample of schools that only use within-class ability grouping.

The remaining columns show the OLS and IV estimates for literacy attainment at age 7. While the OLS estimates are positive and significant, the IV estimates are again of opposite sign. The point estimates  $\hat{\beta}_{0,IV}$  are negative but not significantly different from zero: the largest effect size that can be ruled out is 0.031 sd in all schools, and 0.039 sd in schools that only use within-class ability grouping.

### 3. Robustness

We present three robustness checks. The first, reported in columns 1–4 of Table 4, considers alternative specifications for the ability trajectory of a child from birth to school entry ( $P_i$ ): (i) controlling for birthweight, changes in cognitive ability from age 1 to 3 and age 3 to 5, age-at-test at age 5, and changes in non-cognitive ability from age 1 to 3 and age 3 to 5; (ii) controlling for birthweight, changes in cognitive ability from age 1 to 5, age-at-test at age 5, and changes in non-cognitive ability from age 1 to 5. For both specifications and subjects, the results closely mirror our baseline results.

Second, we consider a violation of the exclusion restriction through season-of-birth effects on educational attainment. If children born at certain times of the year are more likely to spend specific ages indoors because of their season of birth, this can lead to differing adult attention or other inputs. These children are then likely to develop different levels of ability.<sup>36</sup> This explanation is typically ruled out because differences between children who are born at the start and end of the academic year are observed across countries in both hemispheres, which adopt different academic year cut-offs. Moreover, Dhuey et al. (2017) counter such arguments by documenting within-family effects of school starting age on long-run outcomes using detailed population-level administrative data from birth records and later in life.

In our context, we further address this concern in two ways. First, we note that there are no significant differences by term of birth in child characteristics, such as child gender or birthweight, and in family background characteristics, such as household income or whether the mother smoked during pregnancy. In short, the samples are well balanced on observables by term of birth. Second, we exploit the fact that parents were asked whether the child was the result of a planned pregnancy (corresponding to 60 per cent of children). We then re-estimate our IV result for the sample of unplanned pregnancies. The result, shown in columns 5 and 6 of Table 4 mirror the baseline findings.

Third, we control for physical traits of the child from birth, that might otherwise correlate to month of birth and be predictors of true ability at age 7.

<sup>36</sup>Buckles and Hungerman, 2013.

TABLE 4  
Robustness checks, second stage

	Mathematics		Literacy		Unplanned pregnancies		Physical controls	
	Alternative specification for prior attainment: $\Delta 1$ to 3, $\Delta 3$ to 5 (1)	$\Delta 1$ to 5 (2)	$\Delta 1$ to 3, $\Delta 3$ to 5 (3)	$\Delta 1$ to 5 (4)	Mathematics (5)	Literacy (6)	Mathematics (7)	Literacy (8)
Assigned to top group	-0.114*** (0.031)	-0.112*** (0.033)	-0.005 (0.026)	-0.002 (0.028)	-0.102** (0.059)	-0.002 (0.048)	-0.074* (0.041)	0.008 (0.030)
F-statistic	16.9	15.0	39.8	32.6	7.41	14.5	16.7	19.6
Ability trajectory	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-at-test	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family background	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	3,006	3,004	3,275	3,004	1,215	1,305	2,759	3,005

Note: \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . Dependent variable: attainment at age 7. Robust standard errors are given in parentheses, and 95 per cent confidence intervals in brackets. The sample is based on children in schools that only use within-class ability grouping. Values show the effect of assignment to the top within-class ability group on attainment at age 7. Attainment is measured by the NFER math test for mathematics and the BAS reading test for literacy. The test scores are normalized so results represent standard deviation effect sizes. All columns report LIML results estimated using a conditional mixed process with term of birth omitted in the second stage. We condition on the following controls throughout: the child's ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in the household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ 's birthweight, cognitive and non-cognitive ability at ages 1, 3 and 5, and their age-at-test for the age 5 cognitive ability measure. Columns 1–4 report slightly alternative specifications for how these ability trajectory controls are constructed. Columns 1 and 3 control for birthweight, changes in cognitive ability from 1 to 3, 3 to 5, age-at-test at age 5, and changes in non-cognitive ability from 1 to 3, 3 to 5. Columns 2 and 4 control for birthweight, changes in cognitive ability from 1 to 5, age-at-test at age 5, and changes in non-cognitive ability from 1 to 5. Columns 5 and 6 restrict the sample to reported unplanned pregnancies. Columns 7 and 8 control for the physical attributes of the child: their height and weight at ages 3, 5 and 7 and their waist measurement at age 7. For each column, we report the  $F$ -statistic on the joint significance of the instrument (terms of birth).



The traits controlled for are height and weight at ages 3, 5 and 7 and waist measurement at age 7. The results in columns 7 and 8 of Table 4 mirror the baseline findings across subjects.

## V. Non-cognitive outcomes

### 1. Set-up and first stage

The MCS allows us to explore the LATE of being assigned high-ability peers on a rich set of non-cognitive outcomes. As these outcomes are child-specific,  $y_i^7$  (not subject-specific), we first need to specify a child-level treatment. Combining information on group assignment across mathematics and literacy, we define a child to be treated as follows: (i) if the school uses within-class grouping for both subjects, the child is assigned to both top groups; (ii) if the school uses within-class grouping for only one subject, the child is assigned to the top group in that subject. As Table 1 showed, the vast majority of schools use grouping for both subjects. There is a high correlation in top-group assignment across subjects: in schools using within-class grouping for both subjects, 81 per cent are assigned to both top groups. We then estimate the following specification:

$$(3) \quad y_i^7 = \beta_0 [D_{ij}^7 \times D_{ik}^7] + \sum_{k=1,3,5} \beta_{1k} P_i^k + \beta_2 A_i^7 + \beta_3 X_i^7 u_{ij}^7,$$

where  $[D_{ij}^7 \times D_{ik}^7] = 1$  if child  $i$  is in both top ability groups ( $j$  and  $k$ ), as defined above. The omitted category consists of children not assigned to both top ability groups.

Table 5 presents first- and second-stage estimates for this combined treatment. Column 1 shows that term of birth remains strongly predictive of assignment to both top groups [ $F > 20$ ]. Relative to those born in the first term, those born in the second term are 7.3 pp less likely to be assigned to both top groups ( $p < 0.01$ ); those born in the last term are 17 pp less likely to be assigned to both top groups ( $p < 0.001$ ). The conditional likelihood to be assigned to both top groups remains significantly different for all three terms of birth, suggesting the instrument continues to satisfy the monotonicity property and shifts pupils from multiple parts of the term-of-birth distribution into the combined top-group assignment.

Columns 2 and 3 of Table 5 show the second-stage IV effects on mathematics and literacy. For both subjects, this reassuringly shows that when using this redefined treatment dummy, we continue to find a negative LATE on attainment. The estimates suggest quasi-random assignment to both groups of high-ability peers, based on teacher assignments induced solely by variation in term of birth, causes attainment to fall by 0.222 sd in mathematics and

TABLE 5  
*Within-class ability grouping, combined top groups*

	First stage (1)	Mathematics (2)	Literacy (3)
Assigned to both top groups		−0.222*** (0.023) [−0.267, −0.177]	−0.098*** (0.026) [−0.148, −0.048]
Born Jan–Apr	−0.073** (0.028)		
Born May–Aug	−0.171*** (0.036)		
<i>F</i> -statistic	20.3		
<i>p</i> -value [Jan–Apr = May–Aug]	0.028		
Equal to OLS estimate ( <i>p</i> -value)		0.000	0.000
Ability trajectory	Yes	Yes	Yes
Age-at-test	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes
Family background	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes
Number of observations	3046	3011	3044

*Note:* \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . The sample is based on children in schools that only use within-class ability grouping. In column 1 (probit regressions, marginal effects), the dependent variable is assignment to both top groups, and the outcome variable is a dummy equal to one if the child is assigned to both top within-class ability groups. We report marginal effects from a probit, of being born in the Spring term (January to April) and Summer term (May to August) on the probability of being assigned to both top within-class ability groups. The omitted category consists of those children born in the Autumn term of the school year (September to December). In columns 2 and 3 (LIML regressions), the dependent variable is attainment at age 7, and the columns show the effect of assignment to both top within-class ability groups on attainment at age 7. Attainment is measured by the NFER math test for mathematics and the BAS reading test for literacy. The test scores are normalized so results represent standard deviation effect sizes. We report limited information maximum likelihood results estimated using a conditional mixed process with term of birth omitted in the second stage. We condition on the following controls throughout: the child's ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in the household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ 's birthweight, cognitive and non-cognitive ability at ages 1, 3 and 5, and their age-at-test for the age 5 cognitive ability measure. For column 1, we report the *F*-statistic on the joint significance of the instrument (terms of birth) and the *p*-value on the null that the marginal effects for the two reported months of birth on assignment to the top group are equal. For columns 2 and 3, we report the *p*-value on the null that the OLS and LIML-IV estimates are equal.

by 0.098 sd in literacy ( $p < 0.001$  for both estimates). The largest effects on attainment we can rule out in 95 per cent confidence intervals are  $-0.177$  sd for mathematics and  $-0.048$  sd for literacy.

## 2. Teacher–child interactions

Table 6 presents results on non-cognitive outcomes. We first consider teacher–child interactions. The MCS teacher survey elicits information on whether the child asks for extra help from the teacher/teaching assistant, and whether the teacher reports that the child has emotional/concentration difficulties affecting learning. Columns 1 and 2 of Panel A show that the OLS effects go in the expected direction: on average, for children assigned to both groups of high-ability peers, their teacher is significantly less likely to report that they receive extra help or have difficulties with school work.

Once group assignment is accounted for, the results are starkly different for the marginal child assigned to both groups of high-ability peers: there is no impact of top-group assignment on help received from the teacher, but such marginally assigned children are significantly more likely to have difficulties with school work. To benchmark the magnitude, we note that in middle/bottom groups, 15 per cent of children are reported to have difficulties. This doubles if they are quasi-randomly assigned to both top groups due to variation in their term of birth, conditional on their ability trajectory from birth to school, age-at-test and other characteristics.

This all further underpins the fact that our exclusion restriction (i.e. the variation exploited in non-productive traits  $v_{it}$  to induce quasi-random assignment to top groups) does not then also lead to the child receiving extra inputs from teachers post-assignment. We further probe this concern over the exclusion restriction by examining the propensity with which a teacher reports providing a child with additional help. Among children marginally assigned to top groups, we find no significant difference in the propensity for teachers to help children born in different months.

## 3. Parental responses

We capture parental responses to within-class group assignment using: (i) an index of the home learning environment provided by parents; (ii) direct involvement by the parents in their child's homework.<sup>37</sup> Columns 3 and 4 of Table 6 reveal that neither OLS estimate differs from zero, but once an

<sup>37</sup>The home learning environment index is based on parental answers to how many times per week they play, sing, read, paint, or go to the library with their child. The homework measure covers parental help with mathematics, literacy or science homework. Both indices are produced using confirmatory factor analysis (where weights are calculated in the factor analysis according to the amount of information in the component).

TABLE 6  
Non-cognitive outcomes

	Teacher–pupil interaction Help from teacher (1)	Difficulties concentrating affect learning (2)	Home learning environment (3)	Parental inputs Help with homework (4)	Child motivation (5)	Child development Socio-emotional skills (6)	Peer relations (7)
<i>Panel A: OLS</i>							
Assigned to both top groups	−0.075*** (0.008)	−0.175*** (0.013)	−0.027 (0.039)	−0.031 (0.050)	0.266*** (0.026)	0.133*** (0.016)	0.143*** (0.030)
<i>Panel B: IV</i>							
Assigned to both top groups	−0.015 (0.132)	0.216** (0.098)	0.160*** (0.043)	0.107 (0.113)	0.301*** (0.068)	−0.036 (0.046)	−0.433*** (0.010)
Mean outcome in M/B groups	0.132	0.153	0.010	0.015	−0.123	−0.144	−0.080
Ability trajectory	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-at-test	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of obs. (mathematics)	4609	4705	2059	1413	4386	4712	4653

*Note:* \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . Robust standard errors in parentheses. The sample is based on children in schools that only use within-class ability grouping. Values show the effect of assignment to both top within-class ability groups on various mechanisms. Columns 1 and 2 show, respectively, the effect on teacher-reported indicators of whether the child asks for extra help from the teacher or teaching assistant, or whether the child has emotional and concentration difficulties that affect their learning. Columns 3 and 4 show the effect of self-reported parental investments (index of home learning environment metrics) and help with homework, respectively. Column 5 shows the effect on a composite index of child motivations (including self-reported indicators of ‘likes school’, ‘likes numbers’, ‘likes reading’, ‘teacher thinks I am clever’ and ‘trying hard’). Column 6 shows effects on socio-emotional skills at age 7 (measured using an index from the SDQ). Column 7 shows the effect on ‘peer relations’, a composite index of teacher-reported metrics of social interactions including bullying, sharing and tantrums. Panel A (not accounting for group assignment) presents OLS results for mathematics and literacy. Panel B (accounting for group assignment) reports LJM results estimated using a conditional mixed process with term of birth omitted in the second stage. Attainment is measured by the NFER math test for mathematics and the BAS reading test for literacy. The test scores are normalized so results represent standard deviation effect sizes. We condition on the following controls throughout: the child’s ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in the household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ ’s birthweight, cognitive and non-cognitive ability at ages 1, 3 and 5, and their age-at-test for the age 5 cognitive ability measure.

endogenous set assignment is accounted for, parents respond to their child being marginally assigned to high-ability peers by improving the home learning environment (by 0.16 sd relative to children assigned to lower-ability peers). This suggests parents might respond to the negative effects on cognitive outcomes from top-group assignment documented earlier (and more so than teachers). More broadly, the complementarity between classroom practices and parental behaviour dovetails neatly with a literature examining parental responses to educational inputs/school quality.<sup>38</sup>

#### 4. Child development

Given the physical proximity of ability groups in the same classroom, assignment to groups is visible and potentially salient to children. We thus analyse a third dimension of non-cognitive outcomes: child motivation and socio-emotional development.

Children report on multiple attitudes towards school. We combine these into one motivation index (where higher values correspond to being more motivated in school/to study). Column 5 shows that OLS and IV estimates are both of the expected sign: children assigned to both top groups have significantly higher motivation. Digging deeper into components of the motivation index, Table 7 shows that the overall effects for both OLS and IV results are driven by children liking school and reporting the ‘teacher thinks I am clever’.

The socio-emotional development index is as described earlier: higher values denote that the child is psychologically more mature or developed. Column 6 in Table 6 shows the result: again, the OLS estimate is positive and significant, but the IV estimate suggests no effect of marginal assignment to high-ability peers on children’s socio-emotional development in early years. However, this masks important heterogeneity across dimensions of development. Table 8 breaks down the components of the socio-emotional development index. Although the OLS estimates are positive and significant for all components, the IV estimates are more muted and reveal two offsetting effects for the marginal child: a positive effect of high-ability peers on externalizing, but a negative effect on emotional development. These results dovetail neatly with the next non-cognitive outcome we consider: peer relations.

#### 5. Relations with classroom peers

Teachers report on each child’s relations with their classroom peers (not necessarily those in the same ability group). We aggregate these into an overall

<sup>38</sup>Das et al., 2013; Pop-Eleches and Urquiola, 2013; Frederiksson, Ockert and Oosterbeek, 2016; Greaves et al., 2019.

TABLE 7  
*Child motivation components*

	Likes school	Likes numbers	Likes reading	'Teacher thinks I am clever'	Tries hard
	(1)	(2)	(3)	(4)	(5)
<i>Panel A. OLS</i>					
Assigned to both top groups	0.087*** (0.018)	0.062*** (0.018)	0.074*** (0.018)	0.107*** (0.018)	0.060*** (0.014)
<i>Panel B. IV</i>					
Assigned to both top groups	0.230** (0.109)	0.129 (0.116)	0.169 (0.139)	0.299*** (0.113)	0.141 (0.094)
Mean outcome in M/B groups	0.484	0.518	0.541	0.430	0.775
Ability trajectory	Yes	Yes	Yes	Yes	Yes
Age-at-test	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes
Family characteristics	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes
Number of observations	4,479	4,468	4,472	4,425	4,476

*Note:* \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . Robust standard errors are given in parentheses. The sample is based on children in schools that only use within-class ability grouping. Values show the effect of assignment to both top within-class ability groups on various mechanisms related to child motivation. Panel A (not accounting for group assignment) presents OLS results for math and literacy respectively. Panel B (accounting for group assignment) reports LIML results estimated using a Conditional Mixed Process with term of birth omitted in the second stage. Attainment is measured by the NFER math test for math and the BAS reading test for literacy. The test scores are normalized so results represent standard deviation effect sizes. We condition on the following controls throughout: the child's ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ 's birthweight, cognitive and non-cognitive ability at ages one, three and five, and their age-at-test for the age 5 cognitive ability measure.

index (where higher values denote better peer relations). Column 7 in Table 6 shows the results. The OLS estimate in Panel A is positive, implying that, on average, children assigned to both top ability groups have significantly better peer relations. The IV estimate shows a significantly negative effect on the marginal children's relations with classroom peers.

Table 9 digs deeper into the index components. OLS estimates show better peer relations in nearly all components, but the IV estimates are all in the opposite sign and show significantly worsening peer relations on all seven components. For example, the marginal child assigned to both sets of high-ability peers, due to variation in their month of birth, is significantly less likely

TABLE 8  
*Socio-emotional skills components*

	Independence (1)	Prosocial (2)	Externalizing (3)	Emotional (4)
<i>Panel A. OLS</i>				
Assigned to both top groups	0.253*** (0.018)	0.085*** (0.015)	0.036** (0.018)	0.042*** (0.011)
<i>Panel B. IV</i>				
Assigned to both top groups	-0.035 (0.031)	-0.042 (0.071)	0.021*** (0.006)	-0.112*** (0.018)
Mean outcome in M/B groups	-0.217	-0.123	-0.075	-0.078
Ability trajectory	Yes	Yes	Yes	Yes
Age-at-test	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes
Family characteristics	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes
Number of observations	4,712	4,712	4,712	4,712

Notes: \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . Robust standard errors are given in parentheses. The sample is based on children in schools that only use within-class ability grouping. Values show the effect of assignment to both top within-class ability groups on various mechanisms related to the child's socio-emotional skills. Panel A (not accounting for group assignment) presents OLS results for mathematics and literacy. Panel B (accounting for group assignment) reports LIML results estimated using a conditional mixed process with term of birth omitted in the second stage. Attainment is measured by the NFER math test for mathematics and the BAS reading test for literacy. The test scores are normalized so results represent standard deviation effect sizes. We condition on the following controls throughout: the child's ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in the household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ 's birthweight, cognitive and non-cognitive ability at ages 1, 3 and 5, and their age-at-test for the age 5 cognitive ability measure.

to share, more likely to have tantrums, likely to be solitary, and likely to be bullied and fight with others.

These results tie in with the earlier findings on the dimensions of socio-emotional development affected by top-group assignment: the IV estimates in Table 8 showed positive effects of top-group assignment on externalizing, but a negative effect on emotional development.

Given that most mechanisms for relative age effects suggest older children in a class should be more self-confident, etc., these IV estimates for peer relations suggest that the exclusion restriction is unlikely to be violated because the instrument picks up relative age effects.



TABLE 9  
Peer relations components

	Shares well (1)	Has tantrums (2)	Solitary (3)	At least one friend (4)	Bullied (5)	Is well liked (6)	Bullies and fights with other children (7)
<i>Panel A. OLS</i>							
Assigned to both top groups	0.010 (0.006)	-0.016 (0.010)	-0.052*** (0.014)	0.014** (0.006)	-0.054*** (0.011)	0.019*** (0.004)	-0.024*** (0.008)
<i>Panel B. IV</i>							
Assigned to both top groups	-0.065*** (0.003)	0.167*** (0.004)	0.217*** (0.004)	-0.069*** (0.003)	0.179*** (0.004)	-0.045*** (0.004)	0.145*** (0.004)
Mean outcome in M/B groups	0.955	0.113	0.207	0.948	0.148	0.968	0.085
Ability trajectory	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-at-test	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
School characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	4,700	4,705	4,700	4,698	4,703	4,697	4,694

Note: \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ . Robust standard errors are given in parentheses. The sample is based on children in schools that only use within-class ability grouping. Values show the effect of assignment to both top within-class ability groups on various mechanisms related to peer relations. Panel A (not accounting for group assignment) presents OLS results for mathematics and literacy. Panel B (accounting for group assignment) reports LIML results estimated using a conditional mixed process with term of birth omitted in the second stage. Attainment is measured by the NFER math test for mathematics and the BAS reading test for literacy. The test scores are normalized so results represent standard deviation effect sizes. We condition on the following controls throughout: the child's ability trajectory from birth to school entry, their age-at-test for the age 7 tests, binary indicators for white ethnicity and gender, whether living in England, living in London, whether the mother left school between 16 and 18, whether the mother has post-18 education, whether the mother is working, whether the father is present in the household, the log of family net income, the number of siblings, and binary indicators for class size greater than 25, multiple classes per school year, and mixed year group classes. The ability trajectory measures include child  $i$ 's birthweight, cognitive and non-cognitive ability at ages 1, 3 and 5, and their age-at-test for the age 5 cognitive ability measure.

## VI. Discussion

We provide a novel contribution to the long-standing literature on the effects of high-ability classroom peers. We do so by using rich panel data on a representative sample of children in 4,000 primary schools, tracked from birth into early years. We combine this with an IV-based research design that exploits the quasi-random variation in a child's within-class ability group assignment driven by their month of birth. Our key result is that for the marginal child assigned to higher-ability peers, there are no significantly positive effects on cognitive achievement at age 7, and the effects on a number of important non-cognitive outcomes are negative: these include children's motivation towards school/study, difficulties faced with school work, and relations with classroom peers. A clear policy implication is for teachers and schools to try and insulate marginally assigned children from these detrimental effects of being assigned to the top group, say through targeted help from teachers and helping them maintain good relations with classroom peers.

We conclude by discussing: (i) students' concerns about rank; (ii) within-class ability grouping with respect to tracking; (iii) long-run effects of grouping in early years.

First, throughout our analysis we interpret the causal effects on attainment and mechanisms as being driven by a teacher's quasi-random assignment of children to groups. A nascent literature identifies rank concerns as micro-foundations for non-linear peer effects in classrooms.<sup>39</sup> The documented detrimental effects on the marginal child could be driven by rank concerns (where those ranked last in a top-ability group are negatively affected relative to the alternative of being assigned to the top of a lower-ability group).

We still view this interpretation as representing a misallocation of children to groups by teachers, who fail to take account of the rank concerns of children (indeed, such rank concerns can exist in the absence of misassignment). An avenue for future research is to structurally estimate whether rank concerns best explain attainment outcomes. However, to do so requires data that combine the strength of the MCS panel data following children from birth, with the strength of administrative records that cover all children in a class over which rank concerns exist.

Second, we emphasize that our LATE estimates apply to the marginal child assigned to the top group. We cannot extrapolate the effects of high-ability peers on inframarginal always-taker children, who are assigned to the top within-class group irrespective of their month of birth: depending on the nature of peer effects, such children might have large positive effects from high-ability peers. We also reiterate that our research design avoids making cross-school comparisons. Hence, we make no claim as to whether the marginal child does

<sup>39</sup>Tincani, 2017; Cicala, Fryer and Spenkuch, 2018; Murphy and Weinhardt, 2018.

better or worse relative to being in a school with no ability grouping policy at all. Among the most credible evidence for such across-school policy claims comes from Duflo et al. (2011) who implemented a randomized controlled trial on tracking in 121 elementary schools in Kenya. Treated schools were provided with a new teacher, and for these teachers, students were randomly assigned to tracked and non-tracked groups. They documented a significant increase in test scores of 0.175 sd in tracked schools, with achievement effects persisting. All ability groups benefit from tracking, and a key driver of changes is the teachers' improved practices. Of course, Duflo et al. (2011) stress that their results might not be externally valid to high-income settings because their average class size is 46, new teachers were used in the treatment, a key mechanism driving the effects was the teacher's actual class attendance, and the experiment was known to be short term so there was no full adjustment of lesson plans. It remains a first-order priority for this kind of study to be replicated in high-income contexts to understand the effects of within-class ability grouping and school-level tracking jointly.

Finally, our findings show that the use of within-class ability grouping has significant effects on non-cognitive outcomes for the marginal child. This suggests that future research should widen the outcomes considered when studying school-based inputs/practices, moving beyond a focus on cognitive outcomes.<sup>40,41</sup> This remains a policy concern in the UK, which performs poorly in international comparisons of mental health of young people in rich economies, especially in education-related child well-being.<sup>42</sup>

We document all these effects for children in early years, when they are ability grouped between school entry and age 7. This has been shown to be a critical time for the production of human capital, with skills accumulated being complementary to later learning.<sup>43</sup> Our results open up a broad new agenda to study whether such practices in early years drive the formation of cognitive and non-cognitive skills, and this in turn has far-reaching impacts in later stages of the education system and labour market transitions.

<sup>40</sup>Fabregas, 2017; Jackson, 2018.

<sup>41</sup>Fabregas (2017) uses a regression discontinuity design and administrative data from Mexico to show that attending better and more selective middle schools causes children to have lower grade point averages and to perform worse on non-standardized school-based assessments. She further provides evidence to suggest that students evaluate themselves based on their relative performance: marginally admitted students report feeling academically inferior to their peers, have lower self-reported perseverance and time-management scores, and are more likely to shift their aspirations and later schooling choices from academic to vocational programmes. Jackson (2018) uses administrative data on all public school ninth-graders in North Carolina from 2005 to 2012 in order to simultaneously study the effects of teachers on test scores and behaviours (i.e. a proxy for non-cognitive skills). He finds that teachers have meaningful effects on both skill dimensions, but that teacher effects on test scores and on behaviours are only weakly correlated.

<sup>42</sup>UNICEF, 2013.

<sup>43</sup>Cunha et al., 2010.

There is a body of literature that documents the persistent impacts of school starting age across OECD countries on later life outcomes.<sup>44</sup> Some of this might be driven by parents delaying their child's school entry. However, the effect that ability grouping has on the formation of non-cognitive skills in early years might also play an important role in driving long-run outcomes.<sup>45</sup> Such questions are especially relevant when considering grouping practices, given concern over the persistence in group assignment, leading to a misallocation of students to groups.<sup>46</sup> Moreover, many cross-country studies support the idea that family background is more important in countries that group or track students at an early age,<sup>47</sup> suggesting that school-based grouping policies in early years might even lead to persistent disadvantage across generations.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

- Online Appendix

## References

- Abdulkadiroglu, A., Angrist, J. D. and Pathak, P. A. (2014), 'The elite illusion: achievement effects at Boston and New York Exam Schools', *Econometrica*, vol. 82, pp. 137–96.
- Ammermueller, A. and Pischke, J.-S. (2006), 'Peer effects in European primary schools: evidence from PIRLS', National Bureau of Economic Research (NBER) Working Paper no. 12180.
- Angrist, J. and Lang, K. (2004), 'Does school integration generate peer effects? Evidence from Boston's Metco Program', *American Economic Review*, vol. 94, no. 5, pp. 1613–34.
- Argys, L. M., Rees, D. I. and Brewer, D. J. (1996), 'Detracking America's schools: equity at zero cost?', *Journal of Policy Analysis and Management*, vol. 15, pp. 623–45.
- Barker Lunn, J. C. (1970), *Streaming in the Primary School*, Slough: National Foundation for Educational Research in England and Wales.
- Bedard, K. and Dhuey, E. (2006), 'The persistence of early childhood maturity: international evidence of long-run age effects', *Quarterly Journal of Economics*, vol. 121, pp. 1437–72.
- and — (2012), 'School-entry policies and skill accumulation across directly and indirectly affected individuals', *Journal of Human Resources*, vol. 47, pp. 643–83.
- Betts, J. R. (2011), 'The economics of tracking in education', in E. Hanushek, S. Machin and L. Wößmann (eds), *Handbook of the Economics of Education*, Vol. 3, Amsterdam: North-Holland, pp. 341–81.
- and Shkolnik, J. L. (2000a), 'The effects of ability grouping on student achievement and resource allocation in secondary schools', *Economics of Education Review*, vol. 19, 1–15.

<sup>44</sup>Bedard and Dhuey, 2006, 2012; Fredriksson and Ockert, 2014; Dhuey et al., 2017.

<sup>45</sup>Ludwig and Miller, 2007; Chetty et al., 2011; Cornelissen and Dustmann, 2019.

<sup>46</sup>Murphy and Weinhardt, 2018.

<sup>47</sup>Betts, 2011.

- and — (2000b), ‘Key difficulties in identifying the effects of ability grouping on student achievement’, *Economics of Education Review*, vol. 19, pp. 21–26.
- Black, S., Devereux, P. and Salvanes, K. (2007), ‘From the cradle to the labor market? The effect of birth weight on adult outcomes’, *Quarterly Journal of Economics*, vol. 122, pp. 409–39.
- , — and — (2011), ‘Too young to leave the nest? The effects of school starting age’, *Review of Economics and Statistics*, vol. 93, pp. 455–67.
- Booij, A., Leuven, E. and Oosterbeek, H. (2017), ‘Ability peer effects in university: evidence from a randomized experiment’, *Review of Economic Studies*, vol. 84, pp. 547–78.
- Bound, J., Jaeger, D. and Baker, R. (1995), ‘Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak’, *Journal of the American Statistical Association*, vol. 90, pp. 443–50.
- Buckles, K. and Hungerman, D. (2013), ‘Season of birth and later outcomes: old questions, new answers’, *Review of Economics and Statistics*, vol. 95, pp. 711–24.
- Card, D. and Giuliano, L. (2016), ‘Can tracking raise the test scores of high-ability minority students?’, *American Economic Review*, vol. 106, no. 10, pp. 2783–816.
- Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Whitmore Schanzenbach, D. and Yagan, D. (2011), ‘How does your kindergarten classroom affect your earnings? Evidence from Project STAR’, *Quarterly Journal of Economics*, vol. 126, pp. 1593–660.
- , — and Rockoff, J. (2014), ‘Measuring the impact of teachers II: Teacher value-added and student outcomes in adulthood’, *American Economic Review*, vol. 104, no. 9, pp. 2633–79.
- Cicala, S., Fryer, R. G. Jr and Spenkuch, J. L. (2018), ‘Self-selection and comparative advantage in social interactions’, *Journal of the European Economic Association*, vol. 16, pp. 983–1020.
- Clark, D. and Del Bono, E. (2016), ‘The long-run effects of attending an elite school: evidence from the United Kingdom’, *American Economic Journal: Applied Economics*, vol. 8, pp. 150–76.
- Cornelissen, T. and Dustmann, C. (2019), ‘Early school exposure, test scores, and noncognitive outcomes’, *American Economic Journal: Economic Policy*, vol. 11, no. 2, pp. 35–63.
- Crawford, C., Dearden, L. and Greaves, E. (2014), ‘The drivers of month-of-birth differences in children’s cognitive and non-cognitive skills’, *Journal of the Royal Statistical Society, Series A*, vol. 177, pp. 829–60.
- Cullen, J. B., Jacob, B. A. and Levitt, S. (2006), ‘The effect of school choice on participants: evidence from randomized lotteries’, *Econometrica*, vol. 74, pp. 1191–230.
- Cunha, F., Heckman, J. J. and Schennach, S. M. (2010), ‘Estimating the elasticity of intertemporal substitution in the formation of cognitive and non-cognitive skills’, *Econometrica*, vol. 78, pp. 883–931.
- Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. and Sundararaman, V. (2013), ‘School Inputs, Household Substitution, and Test Scores’, *American Economic Journal: Applied Economics*, vol. 5, pp. 29–57.
- Dieterle, S. G., Guarino, C., Reckase, M. and Wooldridge, J. M. (2015), ‘How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added’, *Journal of Policy Analysis and Management*, vol. 34, pp. 461–80.
- Ding, W. and Lehrer, S. F. (2007), ‘Do peers affect student achievement in China’s secondary schools?’, *Review of Economics and Statistics*, vol. 89, pp. 300–12.
- Dobbie, W. and Fryer, R. G. Jr (2011), ‘Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children’s Zone’, *American Economic Journal: Applied Economics*, vol. 3, pp. 158–87.
- Dhuey, E., Figlio, D., Karbownik, K. and Roth, J. (2017), ‘School starting age and cognitive development’, National Bureau of Economic Research (NBER) Working Paper no. 23660.

- Duflo, E., Dupas, P. and Kremer, M. (2011), 'Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya', *American Economic Review*, vol. 101, no. 5, pp. 1739–74.
- Epple, D. and Romano, R. (2011), 'Peer effects in education: a survey of the theory and evidence', in J. Benhabib, A. Bisin and M. Jackson (eds), *Handbook of Social Economics*, Vol. 1B, Amsterdam: North-Holland, pp. 1053–63.
- Fabregas, R. (2017), 'A better school but a worse position? The effects of marginal middle school admissions in Mexico City', mimeo, Harvard.
- Figlio, D. N. and Page, M. E. (2002), 'School choice and the distributional effects of ability tracking: does separation increase inequality?', *Journal of Urban Economics*, vol. 51, pp. 497–514.
- , Guryan, J., Karbownik, K. and Roth, J. (2014), 'The effects of poor neonatal health on children's cognitive development', *American Economic Review*, vol. 104, no. 12, pp. 3921–55.
- Fredriksson, P. and Ockert, B. (2014), 'Life-cycle effects of age at school start', *Economic Journal*, vol. 124, pp. 977–1004.
- , Ockert, B. and Oosterbeek, H. (2016), 'Parental responses to public investments in children: evidence from a maximum class size rule', *Journal of Human Resources*, vol. 51, pp. 832–68.
- Garlick, R. (2018), 'Academic peer effects with different group assignment rules: residential tracking versus random assignment', *American Economic Journal: Applied Economics*, vol. 10, no. 3, pp. 345–69.
- Greaves, E., Hussain, I., Rabe, B. and Rasul, I. (2019), 'Parental responses to information about school quality: evidence from linked survey and administrative data', mimeo UCL.
- Hallam, S. and Parsons, S. (2013), 'Prevalence of Streaming in UK Primary Schools: Evidence from the Millennium Cohort Study', *British Educational Research Journal*, vol. 39, pp. 514–44.
- Hanushek, E., Kain, J., Markman, J. and Rivkin, S. (2003), 'Does peer ability affect student achievement?', *Journal of Applied Econometrics*, vol. 18, pp. 527–44.
- and Wößmann, L. (2006), 'Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries', *Economic Journal*, vol. 116, pp. C63–76.
- Hoekstra, M. (2009), 'The effect of attending the flagship state university on earnings: a discontinuity-based approach', *Review of Economics and Statistics*, vol. 91, pp. 717–24.
- Hoffer, T. B. (1992), 'Middle school ability grouping and student achievement in science and mathematics', *Educational Evaluation and Policy Analysis*, vol. 14, pp. 205–27.
- Hoxby, C. (2000), 'Peer effects in the classroom: learning from gender and race variation', National Bureau of Economic Research (NBER) Working Paper no. 7867.
- Jackson, C. K. (2010), 'Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago', *Economic Journal*, vol. 120, pp. 1399–429.
- (2018), 'What do test scores miss? The importance of teacher effects on non-test score outcomes', *Journal of Political Economy*, vol. 126, pp. 2072–107.
- Krueger, A. B. (1999), 'Experimental estimates of education production functions', *Quarterly Journal of Economics*, vol. 114, pp. 497–532.
- Lefgren, L. (2004), 'Educational peer effects and the Chicago public schools', *Journal of Urban Economics*, vol. 56, pp. 169–91.
- Loveless, T. (2013), '2013 Brown Center Report on American Education: How well are American students learning?', Washington DC: Brookings Institution.
- Ludwig, J. and Miller, D. L. (2007), 'Does Head Start improve children's life chances? Evidence from a regression discontinuity design', *Quarterly Journal of Economics*, vol. 122, pp. 159–208.

- Meghir, C. and Palme, M. (2005), 'Educational reform, ability, and family background', *American Economic Review*, vol. 95, no. 1, pp. 414–24.
- Murphy, R. and Weinhardt, F. (2018), 'Top of the class: the importance of ordinal rank', National Bureau of Economic Research (NBER) Working Paper no. 24958 (*Review of Economic Studies*, forthcoming).
- Neidell, M. and Waldfogel, J. (2010), 'Cognitive and Noncognitive Peer Effects in Early Education', *Review of Economics and Statistics*, vol. 92, pp. 562–76.
- Oakes, J. (2005), *Keeping Track: How Schools Structure Inequality*, 2nd edn, New Haven: Yale University Press.
- Parsons, S. and Hallam, S. (2014), 'The impact of streaming on attainment at age seven: evidence from the Millennium Cohort Study', *Oxford Review of Education*, vol. 40, pp. 567–89.
- Pischke, J.-S. and Manning, A. (2006), 'Comprehensive versus selective schooling in England and Wales: what do we know?', National Bureau of Economic Research (NBER) Working Paper no. 12176.
- Pop-Eleches, C. and Urquiola, M. (2013), 'Going to a better school: effects and behavioral responses', *American Economic Review*, vol. 103, no. 4, pp. 1289–324.
- Rees, D. I., Brewer, D. J. and Argys, L. M. (2000), 'How should we measure the effects of ability grouping on student performance?', *Economics of Education Review*, vol. 19, pp. 17–20.
- Rosenbaum, J. E. (1980), 'Track misperceptions and frustrated college plans: an analysis of the effects of tracks and track misperceptions in the National Longitudinal Survey', *Sociology of Education*, vol. 53, pp. 74–88.
- Slavin, R. E. (1990), 'Achievement effects of ability grouping in secondary schools: a best-evidence synthesis', *Review of Educational Research*, vol. 60, pp. 471–99.
- Tincani, M. (2017), 'Heterogeneous peer effects and rank concerns: theory and evidence', CESifo Working Paper Series no. 6331.
- UNICEF (2013), 'Child well-being in rich countries: a comparative overview', UNICEF Office of Research Innocenti Report Card 11.
- Vardardottir, A. (2013), 'Peer effects and academic achievement: a regression discontinuity approach', *Economics of Education Review*, vol. 36, pp. 108–21.
- Zimmer, R. (2003), 'A new twist in the educational tracking debate', *Economics of Education Review*, vol. 22, pp. 307–15.