



# In-class 'ability'-grouping, teacher judgements and children's mathematics self-concept: evidence from primary-aged girls and boys in the UK Millennium Cohort Study

Tammy Campbell

**To cite this article:** Tammy Campbell (2021) In-class 'ability'-grouping, teacher judgements and children's mathematics self-concept: evidence from primary-aged girls and boys in the UK Millennium Cohort Study, Cambridge Journal of Education, 51:5, 563-587, DOI: [10.1080/0305764X.2021.1877619](https://doi.org/10.1080/0305764X.2021.1877619)

**To link to this article:** <https://doi.org/10.1080/0305764X.2021.1877619>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 19 Mar 2021.



[Submit your article to this journal](#)



Article views: 6939



[View related articles](#)



[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)



# In-class ‘ability’-grouping, teacher judgements and children’s mathematics self-concept: evidence from primary-aged girls and boys in the UK Millennium Cohort Study

Tammy Campbell

Centre for Analysis of Social Exclusion, London School of Economics and Political Science, London, UK

## ABSTRACT

This paper analyses English Millennium Cohort Study data (N = 4463). It examines two respective predictors of children’s maths self-concept at age 11: earlier in-class maths ‘ability’ group and earlier teacher judgements of children’s maths ‘ability/attainment’ (both at age seven). It also investigates differential associations by maths cognitive test score at seven (which proxies maths skill), and by gender. In the sample overall, controlling for numerous potential confounders including maths score, bottom-grouped children and children judged ‘below average’ are much more likely to have later negative maths self-concept. Beneath this aggregate lies variation by gender. All highest ‘ability’-grouped boys have very low chances of negative self-concept, regardless of maths score – but low-scoring girls placed in the highest group have heightened chances of thinking subsequently they are not good at maths. Additionally, the association between negative teacher judgement and negative self-concept is more pervasive for girls.

## ARTICLE HISTORY

Received 24 September 2020

Accepted 12 January 2021

## KEYWORDS

Self-concept; maths; gender; ‘ability’-grouping; teacher judgements; Millennium Cohort Study

## Introduction

Children’s maths self-concept has an impact on their journeys through education and their outcomes beyond. Self-concept can influence learning behaviours, choices of educational tracks and subject specialisms, attainment and adult careers (Hansen & Henderson, 2019; Marsh et al., 2015). Research consistently indicates gendered variation in maths self-concept, with boys tending more often towards a positive view of their own competence, and girls relatively more often to a negative view – a disproportionality not explained by differences in skills (Heyder, Steinmayr, & Kessels, 2019; Sullivan, 2013).

There are known inequalities by gender in outcomes related to maths self-concept, with underrepresentation of girls and women in Science, Technology, Engineering and Maths (STEM) subjects and careers (Codioli McMaster, 2017; Lazarides & Lauermann, 2019). Boaler (1997) argues: ‘If we are to understand the reasons for the underachievement of girls it must surely be necessary to interpret their actions within the context of their environment’ (p. 178). Therefore, examining the early classroom and structural factors that may influence maths self-concept, and that might have differential effects for

**CONTACT** Tammy Campbell  [t.campbell1@lse.ac.uk](mailto:t.campbell1@lse.ac.uk)  Centre for Analysis of Social Exclusion, London School of Economics and Political Science, London, UK

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

girls and boys, during the primary years, can contribute to understanding and potentially to tackling lost potential, and inequities by gender. While many factors interplay in self-concept formation, this paper thus considers two feasible candidates within the child's environment: 'ability' grouping and teacher judgements.

### ***'Ability' grouping and teacher judgements as candidate factors that may influence children's maths self-concept during primary school***

#### ***'Ability' grouping***

Research has long indicated that 'ability' grouping shapes children's educational pathways (Francis et al., 2017a; Ireson & Hallam, 1999; Sukhandan & Lee, 1998). The term 'ability grouping' is itself a misnomer: 'ability' groups are not straightforwardly predicated on children's skills. They are decided only partly according to manifest capability, and are stratified according to factors including demographic characteristics, associated social hierarchies and embedded norms. There is error and misallocation in placement, and placement influences children's experiences and progress (Campbell, 2013, 2017; Connolly et al., 2020; Kutnick et al., 2005; Marks, 2014; Parsons & Hallam, 2014).

How then might 'ability' groups shape maths self-concept within this process of stratification and differentiation? There are two potential mechanisms. 'Labelling' effects (Becker, 1963) suggest that children in lower 'ability' groups will have lower self-concept, because they have internalised the norms and messages around the group and its place – and, correspondingly, children in higher groups will have enhanced self-concept. However, alongside and potentially interacting with labelling effects, 'contrast' mechanisms may also play a part, and are exemplified by the 'big-fish-little-pond' theory. This proposes an advantage to more highly skilled children of being placed in an environment with and comparing themselves to children who are, on average, relatively less skilled (Marsh et al., 2018). Possibly, then, being situated in a lower maths 'ability' group could result for some children in relatively elevated maths self-concept – with corresponding inverse and negative effects of higher group placement.

Recent research examining relationships between maths 'ability' group position and children's maths self-concept, in the UK context, includes Francis et al.'s (2017, 2020) investigations, which collected new empirical data and support the first of these mechanisms: labelling. Francis et al. explore maths set placement at age 11, in the first year of secondary school, and quantify an apparent impact of grouping that widens over time, with bottom-set children having increasingly lowered self-concept, and top-set children heightened. Francis et al. propose an accumulating effect of labelling, denoting self-fulfilling prophecy as 'snowballing prophecy' (2020, p. 14), with 'ability' grouping setting in motion a process that iteratively impacts children's self-concepts via dynamic interactions with their environment and the perceptions and behaviours of those around them, including their teachers.

Qualitative research into the experiences of in-class 'ability'-grouped primary children in the UK correspondingly suggests labelling effects, but also that there are differences across groups in sensitivity or imperviousness to these effects. McGuillicuddy and Devine (2020) report that pupils in their study 'indicated a clear awareness that ability grouping was used because some children are "smarter" than others' (p. 563), and describe how 'being aligned with the high-ability group was a signifier to the children of their higher

ability' (p. 565). However, they also report variations in self-concept and learner identities that intersected with group placement, according to children's characteristics including gender, suggesting that alignment of self-concept with group level does not apply straightforwardly to all pupils. Gripton (2020) studied grouped children in Key Stage One in England, and similarly describes variation in the impacts of the practice that can 'intensify', or be 'mitigate[d]' by, '[t]he scope of the children's awareness' (p. 15).

Other research has also reported ambiguities and nuances beneath the aggregate consequences of 'ability' groupings. For example, Ireson and Hallam (2009) describe how 'different facets of self-concept are sensitive to different aspects of ability grouping in the school as a whole and in specific subjects' (p. 202). Therefore, while at the high level, evidence suggests that 'ability' grouping practices are stratifying, and appear to lead to self-fulfilling (or 'snowballing') prophecies, the totality of their inequitable effects may play out in different ways for different children, through diverse psychological mechanisms, and with varying consequences for children's self-concept.

### *Teacher judgements*

As emphasised by Francis et al.'s (2020) 'snowballing prophecy', one way in which 'ability' grouping has been evidenced to influence children is by the effects of 'labelling' playing out via the perceptions and judgements of teachers. Teachers judge children according to factors including the group in which they are placed (Ansalone, 2003; Boaler, 1997; Boaler, Wiliam, & Brown, 2000; Ireson & Hallam, 1999; Johnston, Wildy, & Shand, 2019). At the same time, interactively, teacher judgements contribute to decisions regarding structuring and placements within 'ability' groupings (Bradbury & Roberts-Holmes, 2017).

Since Rosenthal and Jacobsen's (1968) 'Pygmalion', a literature has built on the impacts of teacher perceptions and judgements, as well as on error and bias in judgements. This includes evidence of a pervasive, disproportionate tendency of teachers to more often rate boys as good at maths, compared to girls (Campbell, 2015; Heyder et al., 2019; Riegle-Crumb & Humphries, 2012; Tiedemann, 2002; Wang, Rubie-Davies, & Meissel, 2018), and indications that judgements to some extent convey individual teachers' own cognitive frameworks and tendencies – rather than simply reflecting children's performance (Rubie-Davies, 2007, 2010).

Heyder et al.'s (2019) recent research into teachers' beliefs suggests that they 'directly affect students' beliefs such as their stereotypes and ability self-concepts', while Timmermans, Rubie-Davies, and Rjosk's (2018) review illustrates that this phenomenon manifests internationally. Correspondingly, analyses of UK national data for the 1958 cohort show that earlier teacher ratings of children's maths 'abilities' predict their later maths self-concept (Sullivan, 2013).

However, as described by Johnston et al. (2019), there is some contention in the literature regarding the substantive significance and relative importance of teacher judgements, and the existence of direct and lasting effects on pupils – including on their self-concept – once other factors, such as classroom structures and children's skills, are taken into account. Jussim and Harber (2005) argue, for example, that their review of '35 years of empirical research' on teacher beliefs shows that '[s]elf-fulfilling prophecies in the classroom do occur, but these effects are typically small ... and they may be more likely to dissipate than accumulate' (p. 131).

## The current study

Firstly, therefore, this paper extends into the primary years the large-scale English quantitative research on maths ‘ability’ grouping and maths self-concept: delineating impacts according to children’s gender and early manifest maths skill, and providing evidence on subgroups potentially differentially impacted by in-class maths ‘ability’ grouping.

Secondly, it adds to estimates of direct and lasting associations between teacher judgements and children’s self-concept, in maths, by looking at longitudinal relationships, in order to disentangle ordering and possible causality – accounting for potential confounders and for corresponding maths ‘ability’ grouping, as well as controlling for and differentiating by gender and measured maths skill level.

Analyses here thus initially explore overall respective associations between both early in-class maths ‘ability’ grouping and early teacher judgements of a child’s maths ‘ability and attainment’ and later maths self-concept, accounting also for whether either of these factors explains the other’s association with self-concept, given their interrelationship and given that the same teacher who provides judgement may have determined in-class groupings. These estimates, for the whole sample, indicate the general importance of each factor in predicting children’s negative maths self-concepts. Then, because maths self-concept varies between girls and boys, and because there is evidence that associations between ‘ability’ groupings and children’s experiences may be heterogeneous, analyses allow variation across children’s manifest maths skills, and by gender.

The main questions addressed are, therefore:

1. Does the maths in-class ‘ability’ group within which a child is placed at age seven predict negative maths self-concept at 11?
2. Does the judgement by their class teacher of a child’s maths ability at age seven predict the child’s negative maths self-concept at 11?
3. Do these relationships vary with a child’s early concurrent maths skill (as measured by maths cognitive test score at age seven)?
4. Do these relationships vary by gender?

## Data

Data is for children, and their teachers and parents, who are taking part in the UK Millennium Cohort Study (MCS), a national longitudinal study of babies born at the turn of the century (<https://cls.ucl.ac.uk/cls-studies/millennium-cohort-study/>). Information from waves three, four and five (ages five, seven and 11)<sup>1</sup> is included. Because education systems and structures vary across UK countries, the sample is restricted to children who attended school in England at age seven (wave four), for whom there are responses to key questions in a survey of their teachers when they were seven, and who have information on maths self-concept at age 11 (wave five). Children who are extremely low-scoring (<6) outliers on the key maths cognitive test variable (N = 38) are removed from the sample to prevent disproportionate influence and skewing of results conditional on the test scores, leaving a total sample of N = 4463. Unless otherwise specified, all main analyses are weighted for the MCS’s stratified, clustered design, and for non-response and attrition to

wave five, using `svy` commands alongside the `subpop` specification, in Stata 14. Because analyses are for a selected sub-sample rather than for the whole wave five sample, unweighted versions of all models are also checked (results are extremely similar).

### **Outcome variable: maths self-concept**

The outcome variable is taken from wave five, when children were 11 years old, and is their response to the self-completion survey question<sup>2</sup>:

‘How much do you agree ... I am good at Maths’.

Children could respond ‘Strongly agree’/‘Agree’/‘Disagree’/‘Strongly disagree’. The variable is recoded as binary, so both ‘agree’ responses are grouped, and both ‘disagree’ responses are combined. As shown in Table 3, most children agree that they are ‘good at maths’; thirteen per cent do not. Analyses examine the odds of children disagreeing to any extent that they are good at maths at age 11 – which is conceptualised as representing negative maths self-concept. A limitation of this work is that the negative self-concept measure thus relies on a single survey item, and measures one facet of self-concept – the child’s perception of their own competence in maths – unlike recent work which incorporates multi-item measures (e.g. Francis et al., 2017, 2020). However, the advantage of the single item approach is clarity and precision of outcome, ease of interpretation and straightforward measurement of children’s reported judgement of their own maths skill.

### **Key predictors**

#### ***Maths ‘ability’ group at age seven***

The MCS children’s teachers were contacted, when children were aged seven,<sup>3</sup> and asked, ‘In this child’s class, are there within-class subject groups for maths?’ and, subsequently, ‘Which group is this child in for maths?’ This results in information that the child is not grouped in-class for maths (17% of the sample), in the highest group (34%), the middle group (35%) or the lowest group (17%). In acknowledgement of the possibility of generalised or cross-domain effects, the equivalent information on group for literacy at seven is also included.<sup>4</sup>

#### ***Teacher judgements of children’s maths ‘ability and attainment’ at age seven***

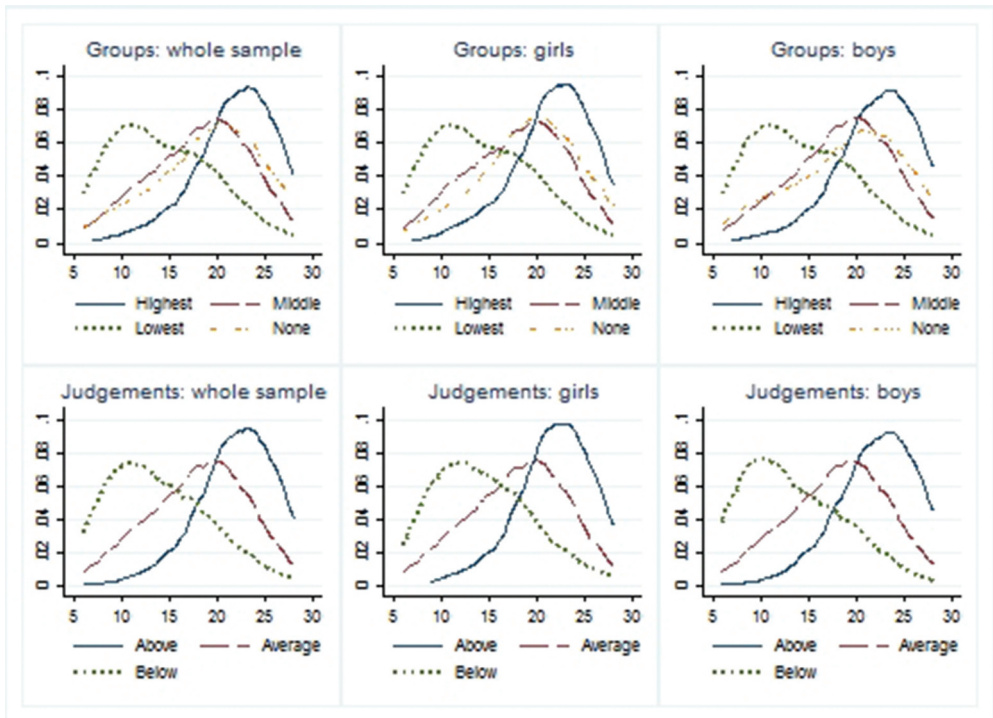
Teachers were additionally asked, when children were seven, to ‘rate the child in relation to all children of this age (i.e. not just their present class or, even, school)’. One domain in which teachers were asked to rate the children was ‘Maths and Numeracy’, and they could respond that the child was ‘Well above average’/‘Above average’/‘Average’/‘Below average’/‘Well below average’. In order to maintain adequate cell sizes, this variable is recoded into three categories, and 43% of the sample’s teachers report them as being above average, 40% as average, and 17% as below average at maths. This represents teachers’ judgements of the children’s maths ability.

Models also incorporate equivalent teacher judgements of children’s reading ability at age seven, again in order to integrate the possibilities both of generalised/domain spill-over or of cross-domain influences. The latter are inverse between-subject relationships evidenced throughout the literature on self-concept:

higher reading competence is related to lower maths self-efficacy (e.g. Chui, 2016; Marsh & Hau, 2004).

### *Maths cognitive test performance at age seven*

Children undertook the NFER Progress in Maths cognitive assessment when they were seven. This test was administered during fieldwork in children's homes (which took place over an approximately six-month-long period before the teacher survey<sup>5</sup>) and 'assesses a child's mathematical skills and knowledge' (Connelly, 2013). The scaled raw score is used; this is transformed to take account of the difficulty levels of test items completed, but not otherwise standardised. By controlling for scores on this test (and for age at test), models examine relationships between early grouping and teacher judgements, and later self-concept, for children who appeared similar in their early concurrent maths skills. As detailed later in this article, maths test score is also interacted, in selected models, with group placement and with teacher judgement, respectively, to examine whether these factors have differential associations with self-concept depending on the manifest skills of the child. Scores for all sample children range from 6 to 28; Figure 1 shows the distributions of scores.



**Figure 1.** Distribution of maths test scores across 'ability' groups and teacher judgements. Whole sample  $N = 4463$ ; girls  $N = 2299$ ; boys  $N = 2164$ . Source: Millennium Cohort Study, wave 4. Horizontal axis = test scores; vertical axis = proportion of children in each group with each score.

## Gender

This is a binary measure based on parent report, and is used as a control in some models, and to separate analyses for girls and boys.

## Controls

An aim of analyses is to determine whether there is an independent relationship between teacher judgement at age seven and, respectively, ability group placement at age seven, and maths self-concept at age 11. Therefore a number of controls that may feasibly precede, account for and influence both earlier groupings and/or judgements, and later self-concept, are included. These span child and family characteristics, scores on other cognitive tests (covering maths, literacy and general domains at ages five and seven), parent judgements and home inputs.<sup>6</sup> [Table 1](#) describes each of the factors, and their raw relationship with maths 'ability' group, while [Table 2](#) does the same for each factor and maths teacher judgement. [Table 3](#) shows the raw relationships between each variable, including maths 'ability' group and maths teacher judgement, and negative maths self-concept at age 11.

In line with previous research on 'ability' grouping among the MCS children (Campbell, 2017, 2013; Hallam & Parsons, 2012), [Table 1](#) shows that those from high-income families are more likely to be in the higher maths 'ability' group, along with those with no teacher-reported special educational needs (SEN), those from families speaking only English at home, those whose mother is educated to degree-level, and those who are relatively older within the school year. Children with higher maths test scores are more likely to be in a higher group, as well as those whose parents report no maths or reading difficulties at seven, and no help with maths or reading at home. Girls are more likely to be in the middle maths 'ability' group and less likely to be in the higher group than boys.

[Table 2](#) shows a similar pattern of relationships with teacher judgements of maths, again in line with previous work using this data (Campbell, 2015). Sample boys are more likely to be judged 'above average', alongside higher-income children, those with no reported SEN, those who speak English only, those with more highly educated mothers, and relatively older children. Children who score higher across all cognitive tests, and, again, those whose parents report no difficulties with maths and reading and no help at home with these subjects, are also more likely to be judged positively at maths by their teacher.

In terms of raw relationships with children's negative maths self-concept at 11, [Table 3](#) shows that those in the lowest maths 'ability' group at age seven are most likely to report not being good at maths at age 11 (25% vs. 5% of those in the highest group). Children who are not in-class grouped for maths have a lower likelihood of later negative self-concept than those placed in the middle group (11% vs. 16%) and compared to the overall average (13%). Children judged 'below average' at maths at seven are also much more likely than those judged 'above average' to have later negative maths self-concept (26% vs. 3%). Children reporting negative maths self-concept at 11 had, on average, lower maths cognitive test scores at seven (mean = 16 vs. mean = 19; range in sample 6–28), and girls are more likely to report not being good at maths at 11 (16%, vs. 9% of boys).



**Table 1.** Raw relationships between each other predictor variable and maths 'ability' group at age seven (highest, middle, lowest).

	Percent/mean			
	Highest (1597)	Middle (1494)	Lowest (598)	None (774)
<i>All sample children</i>	34	35	14	17
<i>Literacy group (age 7)</i>				
Highest (1611)	78	15	1	5
Middle (1493)	18	69	7	6
Lowest (686)	2	26	67	6
None (673)	6	9	2	82
<i>Maths teacher judgement (age 7)</i>				
Above (1974)	67	16	1	16
Average (1739)	14	61	8	17
Below (750)	1	23	62	15
<i>Reading teacher judgement (age 7)</i>				
Above (2236)	56	25	2	17
Average (1430)	19	52	12	17
Below (797)	4	32	48	16
<i>Gender (age 7)</i>				
Girls (2299)	33	37	13	17
Boys (2164)	36	32	15	17
<i>Ethnic group (age 7)</i>				
White (3676)	35	35	14	16
Mixed (141)	37	30	14	20
Indian (151)	26	29	17	27
Pakistani (217)	24	40	13	24
Bangladeshi (65)	36	20	8	35
Black Caribbean (46)	36	31	15	19
Black African (81)	25	38	13	24
Other ethnicity (86)	36	31	12	20
<i>Family income (age 7)</i>				
Higher income (3373)	38	34	12	16
Low income (1090)	24	38	20	17
<i>Teacher-reported Special Educational Needs (age 7)</i>				
Not reported (3592)	40	36	7	17
Ever Special Educational Needs (871)	12	30	42	15
<i>Home language (age 7)</i>				
English only (3986)	35	35	14	16
Additional language(s) (567)	24	38	13	25
<i>Mother's education (age 7)</i>				
Not NVQ 4+ (1853)	31	37	17	16
NVQ 4 or above (2610)	40	32	10	18
<i>Mean month of birth (Sept = 1 – Aug = 12)</i>	5.5	6.6	7.6	6.3
<i>Mean maths score (age 7) – sample range: 6–28</i>	22	18	14	19
<i>Mean reading score (age 7) – range: 10–214</i>	127	107	84	113
<i>Mean nam. vocab. score (age 5) – range: 10–170</i>	115	108	101	109
<i>Mean pic. sim. score (age 5) – range: 10–119</i>	85	81	77	83
<i>Mean pat. const. score (age 5) – range 10–152</i>	97	88	76	90
<i>Parent report of child's maths difficulties (age 7)</i>				
None (3109)	45	32	7	16
Difficulties (1354)	11	42	30	17
<i>Parent report of child's reading difficulties (age 7)</i>				
None (3406)	42	33	8	17
Difficulties (1057)	11	40	35	15
<i>Maths help at home (age 7)</i>				
No help (1810)	38	33	12	17
Help with maths (2653)	31	36	16	16

(Continued)

**Table 1.** (Continued).

	Percent/mean			
<i>Reading help at home (age 7)</i>				
No help (1445)	43	31	9	17
Help with reading (3018)	30	37	17	16

Notes: Ns are unweighted; proportions and means are weighted. Note that though descriptives for these variables are not shown here, all regressions in the results section are controlled for age at respective cognitive test and months lapsed from cognitive test to teacher survey, as appropriate.

Source: Millennium Cohort Study, waves 3 and 4.

**Table 2.** Raw relationship between maths teacher judgement (above average, average, below average) at age seven, and each predictor variable.

	Percent/mean		
	Above average (1974)	Average (1739)	Below average (750)
<i>All sample children</i>	43	40	17
<i>Maths group (age 7)</i>			
Highest (1597)	84	16	1
Middle (1494)	20	69	11
Lowest (598)	2	23	76
Not grouped (774)	43	42	15
<i>Literacy group (age 7)</i>			
Highest (1611)	76	22	2
Middle (1493)	29	57	14
Lowest (686)	5	33	62
Not grouped (673)	41	46	13
<i>Reading teacher judgement (age 7)</i>			
Above (2236)	74	24	2
Average (1430)	19	68	13
Below (797)	4	31	64
<i>Gender (age 7)</i>			
Girls (2299)	39	44	17
Boys (2164)	47	35	18
<i>Ethnic group (age 7)</i>			
White (3676)	44	39	17
Mixed (141)	50	34	16
Indian (151)	45	35	20
Pakistani (217)	28	43	29
Bangladeshi (65)	38	40	22
Black Caribbean (46)	21	56	23
Black African (81)	26	56	18
Other ethnicity (86)	46	33	22
<i>Family income (age 7)</i>			
Higher income (3373)	48	38	14
Low income (1090)	27	45	28
<i>Teacher-reported Special Educational Needs (age 7)</i>			
Not reported (3592)	51	41	9
Ever Special Educational Needs (871)	14	35	51
<i>Home language (age 7)</i>			
English only (3986)	44	39	17
Additional language(s) (567)	35	42	23
<i>Mother's education (age 7)</i>			
Not NVQ 4+ (1853)	36	42	22
NVQ 4 or above (2610)	54	36	10

(Continued)

**Table 2.** (Continued).

	Percent/mean		
	Above average (1974)	Average (1739)	Below average (750)
<i>Mean month of birth (Sept = 1 – Aug = 12)</i>	5.5	6.8	7.4
<i>Mean maths score (age 7) – sample range: 6–28</i>	22	18	14
<i>Mean reading score (age 7) – range: 10–214</i>	127	107	84
<i>Mean nam. vocab score (age 5) – range: 10–170</i>	115	108	100
<i>Mean pic. sim. score (age 5) – range: 10–119</i>	85	81	77
<i>Mean pat. const. score (age 5) – range 10–152</i>	97	87	77
<i>Parent report of child's maths difficulties (age 7)</i>			
None (3109)	55	36	9
Difficulties (1354)	16	48	36
<i>Parent report of child's reading difficulties (age 7)</i>			
None (3406)	52	38	10
Difficulties (1057)	14	45	41
<i>Maths help at home (age 7)</i>			
No help (1810)	50	36	14
Help with maths (2653)	38	42	20
<i>Reading help at home (age 7)</i>			
No help (1445)	54	35	10
Help with reading (3018)	37	41	21

**Table 3.** Raw relationships between each variable and negative maths self-concept at age 11.

All children (N = 4463)	
Percent negative SC	13
<i>'Ability' groups (age 7)</i>	
<i>Maths group: percent negative self-concept</i>	
Highest (1597)	5
Middle (1494)	16
Lowest (598)	25
Not grouped (774)	11
<i>Literacy group: percent negative self-concept</i>	
Highest (1611)	7
Middle (1493)	16
Lowest (686)	20
Not grouped (673)	10
<i>Teacher judgements (age 7)</i>	
<i>Maths teacher judgement: percent negative self-concept</i>	
Above (1974)	3
Average (1739)	15
Below (750)	26
<i>Reading teacher judgement: percent negative self-concept</i>	
Above (2236)	9
Average (1430)	14
Below (797)	19
<i>Child/family characteristics (age 7)</i>	
<i>Gender: percent negative self-concept</i>	
Girls (2299)	16
Boys (2164)	9
<i>Ethnic group: percent negative self-concept</i>	
White (3676)	13
Mixed (141)	10
Indian (151)	9
Pakistani (217)	14

(Continued)

**Table 3.** (Continued).

All children (N = 4463)	
Bangladeshi (65)	11
Black Caribbean (46)	15
Black African (81)	3
Other ethnicity (86)	6
<i>Family income: percent negative self-concept</i>	
Higher income (3373)	12
Low income (1090)	14
<i>Teacher-reported Special Educational Needs: percent negative self-concept</i>	
Not reported (3592)	11
Ever Special Educational Needs (871)	19
<i>Home language: percent negative self-concept</i>	
English only (3986)	13
Additional language(s) (567)	11
<i>Mother's education: percent negative self-concept</i>	
Not NVQ 4+ (1853)	13
NVQ 4 or above (2610)	11
<i>Mean month of birth (Sept = 1 – Aug = 12)</i>	
All children (4463)	6.3
With negative SC (538)	6.5
Without negative SC (3925)	6.3
<i>Cognitive test scores</i>	
<i>Mean maths score (age 7) – range in sample: 6–28</i>	
All children (4463)	19
With negative SC (538)	16
Without negative SC (3925)	19
<i>Mean reading score (age 7) – range in sample: 10–214</i>	
All children (4463)	111
With negative SC (538)	104
Without negative SC (3925)	112
<i>Mean naming vocab score (age 5) – range in sample: 10–170</i>	
All children (4463)	109
With negative SC (538)	107
Without negative SC (3925)	110
<i>Mean picture similarity score (age 5) – range in sample: 10–119</i>	
All children (4463)	82
With negative SC (538)	79
Without negative SC (3925)	83
<i>Mean pattern construction score (age 5) – range in sample 10–152</i>	
All children (4463)	89
With negative SC (538)	84
Without negative SC (3925)	90
<i>Parent judgements and input</i>	
<i>Parent report of child's maths difficulties (age 7): percent negative self-concept</i>	
None (3109)	8
Difficulties (1354)	22
<i>Parent report of child's reading difficulties (age 7): percent negative self-concept</i>	
None (3406)	11
Difficulties (1057)	17
<i>Maths help at home (age 7): percent negative self-concept</i>	
No help (1810)	12
Help with maths (2653)	13
<i>Reading help at home (age 7): percent negative self-concept</i>	
No help (1445)	11
Help with reading (3018)	14

Note: Ns are unweighted, proportions and means are weighted. Note that though descriptives for these variables are not shown here, all regressions in the results section are controlled for age at respective cognitive test and months lapsed from cognitive test to teacher survey, as appropriate.

Source: Millennium Cohort Study, waves, 4 and 5.

**Table 4.** Model specifications (outcome is always negative maths self-concept at 11).

Predictors	Specification											
	1	2	3	4	4a	4b	5	5a	5b	6	6a	6b
Maths 'ability' group at 7	x		x	x	x	x	x	x	x	x	x	x
Maths teacher judgement at 7		x	x	x	x	x	x	x	x	x	x	x
All controls (at 5/7)				x	x	x	x	x	x	x	x	x
Maths 'ability' group * maths test score							x	x	x			
Maths teacher judgement * maths test score										x	x	x
All children	x	x	x	x			x			x		
Girls only					x			x			x	
Boys only						x			x			x
N	4463	4463	4463	4463	2299	2164	4463	2299	2164	4463	2299	2164

## Analytical strategy

Analyses explore relationships between 'ability' group and maths self-concept, and teacher judgement and maths self-concept, accounting for the other factor of interest, as well as the controls detailed in Tables 1–3. Modelling also investigates whether relationships vary according to score at seven on the Progress in Maths cognitive test, and whether there are different patterns for girls and boys.

In order to condition analyses on the maths cognitive test score it is necessary that test scores span children in each 'ability' group and with each level of teacher judgement. Figure 1 shows that this is the case, both in the sample as a whole and when it is divided into girls and boys. While low-scoring children are more likely to be in the lowest 'ability' group and high-scoring children in the highest, it is also the case that children across the range of test scores appear in all groups, with mid-scorers distributed fairly evenly. There is a similar pattern for the distribution of scores by teacher judgement.

Twelve model specifications are used to address the research questions. All are logistic regressions, in which the outcome variable is children's reported negative maths self-concept at 11 (1/0). Table 4 details the predictors included in each specification.

Model-predicted log odds for the key variables (maths group, maths teacher judgement, and test score and gender where included) are reported in tables for each of these regressions, with conversion by exponentiation to odds ratios exemplifying selected findings and discussed in the text. The reference category for maths 'ability' group is set at 'highest', and for maths teacher judgement at 'above average' throughout. Graphs of predicted probabilities estimated for key variables in each model are also presented, to aid interpretation, demonstrate substance and illustrate patterns and relationships.

## Results

Table 5 presents log odds produced by specifications 1–4b. Specification 1 reiterates that sample children placed in the lowest maths 'ability' group at age seven have odds much greater than those placed in the highest group of negative maths self-concept at 11 (log odds: 1.94; OR: 6.97;  $p < 0.001$ ). Specification 2 again corresponds to Table 3's raw figures, showing that sample children judged by their teacher as 'below average' have

higher odds than those judged ‘above average’ of later negative maths self-concept (log odds: 1.87; OR: 6.50;  $p < 0.001$ ).

Specification 3 includes both of these predictors (‘ability’ grouping and teacher judgement) together. In line with previous research indicating their interrelationship, each is attenuated by the other. The predicted odds of a child in the lowest ‘ability’ group having later negative maths self-concept are less starkly contrasted to those of a child in the highest group, once distribution across teacher judgements is taken into account. However, a difference between groups independent of the apparent influence of concurrent teacher judgement remains, with children in the lowest group still estimated to have raised odds compared to those in the highest group (log odds: 0.93; OR: 2.54;  $p < 0.001$ ). Similarly, the relationship between teacher judgement and later self-concept is modified but by no means fully explained by concurrent ‘ability’ group (log odds: 1.31; OR: 3.71;  $p < 0.001$  for children judged ‘below average’ compared to those judged ‘above average’). Thus it seems that both maths in-class ‘ability’ group and teacher judgement of children’s maths at seven have a relationship with later maths self-concept independent of the other.

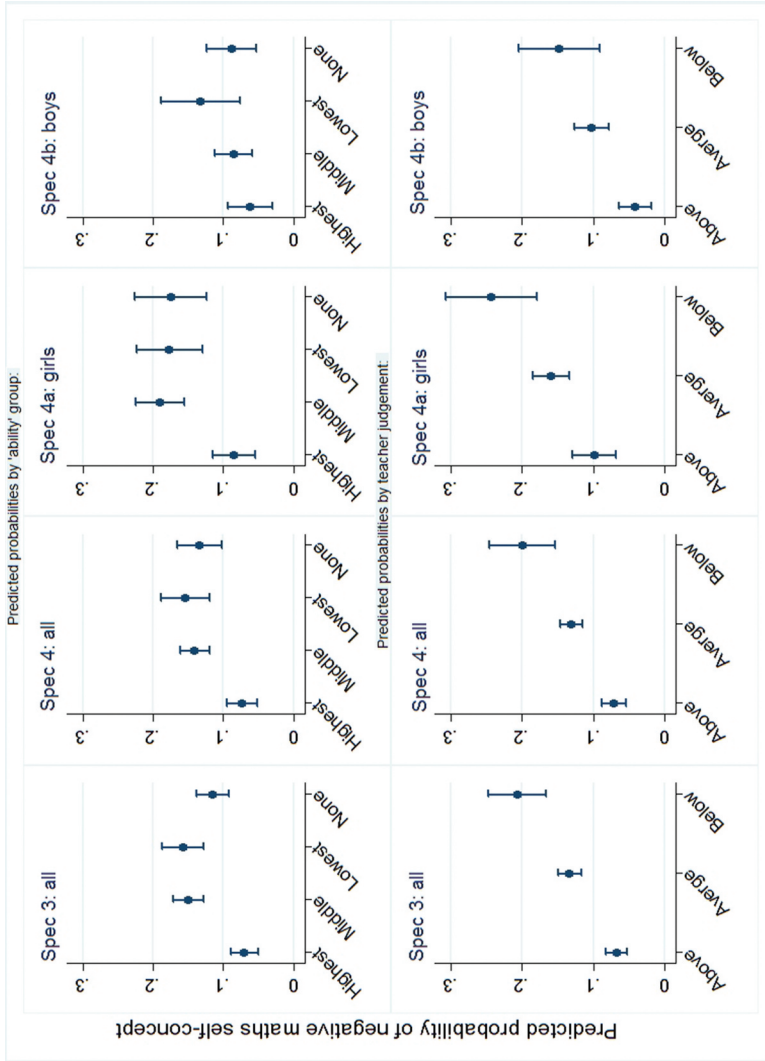
Specification 4 addresses the possibility that third factors may, however, account for these relationships. Controls for maths cognitive test score, child and family characteristics, parent judgements and home input, and other teacher judgements, ‘ability’ groups, and test scores in complementary and contrasting domains are added. Controls including gender and maths test score at seven – as shown in Table 5 – are associated in this model with later maths self-concept (OR for girls is 2.02;  $p < 0.001$ , compared to boys; each maths test score point [range 6–28] is associated with a decrease in odds by 0.96;  $p < 0.001$ ). However, odds ratios for children in the lowest maths group compared to the highest maths group change little on addition of these controls (OR: 2.45;  $p < 0.001$ ); similarly, odds for children judged below average, compared to those judged above average, remained stable (OR: 3.55;  $p < 0.001$ ). Figure 2 illustrates this by showing

**Table 5.** Results – Specifications 1–4b. Relationships of ‘ability’ group placements and teacher judgements with later maths self-concept.

	Spec 1: all	Spec 2: all	Spec 3: all	Spec 4: all	Spec 4a: girls	Spec 4b: boys
Highest maths group (ref)	-	-	-	-	-	-
Middle maths group	1.14***		0.86***	0.78***	0.99***	0.37
Lowest maths group	1.94***		0.93***	0.90**	0.89**	0.91+
No maths group	1.00***		0.55**	0.71**	0.87**	0.41
Above average judgement (ref)		-	-	-	-	-
Average judgement		1.20***	0.77***	0.73***	0.59**	1.02**
Below average judgement		1.87***	1.31***	1.27***	1.15***	1.48***
Boys (ref)				-		
Girls				0.70***		
Maths test score				-0.04**	-0.04**	-0.06**
Controls	No	No	No	Yes	Yes	Yes
N	4463	4463	4463	4463	2299	2164

Notes: \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ ; +  $p < 0.10$ . Table shows log odds. All estimates are weighted for sample design and attrition. As per Table 3, controls are: age at respective cognitive test; months lapsed from cognitive test to teacher survey; literacy ‘ability’ group at 7; reading teacher judgement at 7; ethnicity; family income; Special Educational Needs; home language; mother’s education; month of birth; reading test score at 7; naming vocabulary score at 5; picture similarity score at 5; pattern construction test score at 5; parent report of child’s maths difficulties at 7; parent report of child’s reading difficulties at 7; maths help at home at 7; reading help at home at 7.

Source: Millennium Cohort Study, waves 3, 4 and 5.



**Figure 2.** Predicted probabilities of negative maths self-concept at 11 by 'ability' group or teacher judgement at 7 – Specifications 3–4b. First row = estimated probabilities by group; second row = estimated probabilities by teacher judgement. Specifications 3 and 4 N = 4463; Specification 4a N = 2299; Specification 4b N = 2164. Interpret in conjunction with Table 5. Error bars are 95% CIs.

a continued and substantial difference in model-predicted probabilities of negative self-concept for children in different groups and with different teacher judgements.

These results suggest that, among the sample including both girls and boys, there are independent effects of both maths in-class 'ability' group, and of teacher judgements of children's maths, on children's later maths self-concept. When the sample is divided by gender (Specifications 4a and 4b), both boys and girls in the lowest group are more likely than counterparts of the same gender in the highest group to have negative maths self-concept (OR: 2.49;  $p = 0.05$  for boys; OR: 2.44;  $p = 0.01$  for girls). However, boys in the middle group are no more likely than those in the highest group to have negative self-concept ( $p = 0.30$ ), while girls in the middle group are more likely than girls in the highest group (OR: 2.70;  $p < 0.001$ ).

In Specification 5 (Table 6), maths cognitive test score is interacted with 'ability' group level. There are statistically significant interactions between score and group levels, indicating that relationships between earlier maths skills and later self-concept vary according to the group in which a child is situated. Figure 3 illustrates this with model predicted probabilities for children in the highest and lowest groups, across the range of scores. It suggests a more pronounced relationship between maths skill and later self-concept for those in the highest group, whose lowered odds of negative self-concept are most strongly related to increased maths score (OR: 0.91;  $p < 0.001$ ).

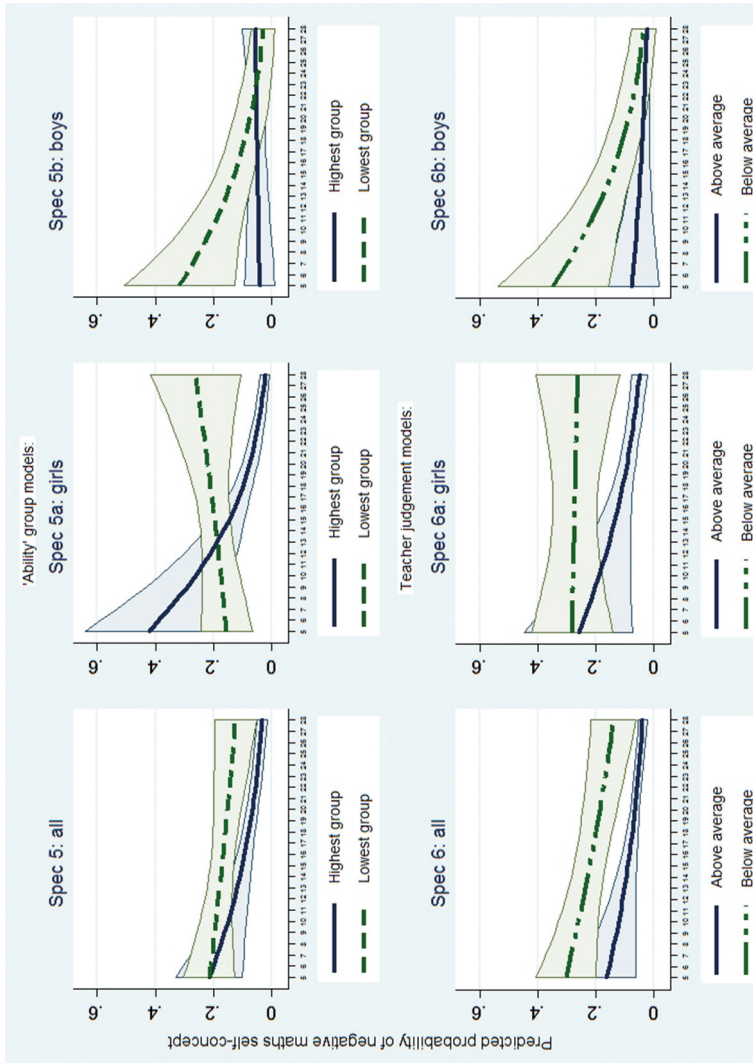
**Table 6.** Results – Specifications 5–6b. Relationships of 'ability' group placements and teacher judgements with later maths self-concept, when each of these factors is interacted with maths cognitive test score.

	Spec 5: all Group*score	Spec 5a: girls Group*score	Spec 5b: boys Group*score	Spec 6: all Judge*score	Spec 6a: girls Judge*score	Spec 6b: boys Judge*score
Highest maths group (ref)	-	-	-	-	-	-
Middle maths group	-0.24	-1.03	1.38	0.76***	0.97***	0.39
Lowest maths group	-0.33	-2.41**	3.31**	0.88**	0.87**	0.92**
No maths group	-0.51	-1.80**	2.12+	0.70**	0.87**	0.46
Middle maths group*Maths score	0.05+	0.10**	-0.05			
Lowest maths group*Maths score	0.06+	0.19***	-0.14**			
No maths group*Maths score	0.08*	0.14**	-0.08			
Above average judgement (ref)	-	-	-	-	-	-
Average judgement	0.70***	0.51**	1.14**	0.06	-0.20	0.37
Below average judgement	1.26***	1.10***	1.54***	0.75	-0.29	2.33**
Average judgement*Maths score				0.03	0.04	0.04
Below average judgement*Maths score				0.02	0.08+	-0.07
Maths test score	-0.09***	-0.15***	0.01	-0.07**	-0.09**	-0.05
Intercept	4.51	5.79	0.54	4.05	4.53	1.81
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	4463	2299	2164	4463	2299	2164

Notes: \*\*\*  $p < 0.001$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ ; +  $p < 0.10$ . Table shows log odds. All estimates are weighted for sample design and attrition. Controls are as per Table 3 and Table 5.

Source: Millennium Cohort Study, waves 3, 4 and 5.





**Figure 3.** Predicted probabilities of negative maths self-concept at 11 – Specifications 5, 5a, 5b (‘ability’ group interaction with maths test score); Specifications 6, 6a, 6b (teacher judgement interaction with maths test score). Specifications 5 and 6a N = 4463; Specification 5a and 6a N = 2299; Specification 5b and 6b N=2164. Interpret in conjunction with Table 6. Shaded areas are 95% CIs around estimate at each value of test score, which is x axis. Y axis is probability of negative maths self-concept.

Once the sample is split into boys and girls, different patterns emerge. For girls (Table 6: Specification 5a), there are significant interactions between maths test score and 'ability' group levels; the model intercept for the girls' lowest group also varies significantly from that for the top group ( $p < 0.01$ ). Figure 3 illustrates the resulting pattern of relationships with predicted probabilities for girls in the highest and lowest groups. While the association between higher score and negative self-concept is negative for girls in the highest 'ability' group, it is significantly different to this ( $p < 0.01$ ) and positive for those in the lowest group. Among higher-scoring girls, high group placement (as opposed to low) is associated with a lower probability of negative maths self-concept, but this is not true for lower-scoring girls. This suggests labelling effects for high-scoring girls, but potential contrast or comparison effects among low-scoring girls, where being placed in a group with relatively more skilled peers, or within which there are higher expectations or norms, may impact negatively on those girls who are currently less skilled, rather than boosting self-concept. Error bars are 95% CIs.

This diverges from a much more straightforward association between high-group placement and boys' self-concept. Specification 5b (Table 6) indicates that the model intercept for boys in the lowest 'ability' group is significantly higher than that for boys in the highest group. At the same time, there is no relationship between maths test score and negative self-concept for high group boys, while there is a negative relationship significantly different from this for boys in the lowest group. As demonstrated by Figure 3, this interaction indicates that skill at seven, as measured by maths test score, is largely unrelated to later self-concept for boys placed in the highest 'ability' group: boys in this group all tend to have a very low probability of subsequent negative self-concept. This supports the possibility of generally positive labelling effects of higher group placement for boys. Low-scoring boys in low groups have a higher probability of saying they are not good at maths, again indicating labelling effects.

Specification 6 (Table 6 and Figure 3) suggests that in the whole sample of girls and boys, the relationships of maths score and teacher judgement with later self-concept do not vary significantly across one another: regardless of judgement level, higher measured maths capability is associated with lower odds of negative self-concept. For girls, however, there is a significant interaction between test score and teacher judgement. As shown in Table 6 and Figure 3, Specification 6a, among girls who are judged 'above average' by their teachers, maths skill is related to self-concept, with high-scoring girls less likely subsequently to view themselves negatively. However, in contrast, across test scores, girls who are judged 'below average' by their teacher at seven are all relatively more likely to have later negative maths self-concept.

### **Sensitivity checks**

Alternative specifications include: testing all interacted models without controls; adding low-scoring outliers back into the sample; using a categorical recoding of the maths score variable, to check for non-linearities; and analyses without survey weights (because the analytical sample is not a complete representation of the wave five sample). All these checks yield results consistent with the main findings.

## Summary and discussion

Returning to the research questions, the results from these analyses of the Millennium Cohort sample children can be summarised as follows.

1. Does the maths in-class ‘ability’ group within which a child is placed at age seven predict negative maths self-concept at 11?

In the sample overall, in-class maths ‘ability’ group at seven predicts maths self-concept at 11, and this association holds at a reduced but still substantial magnitude, both once teacher judgements of maths are accounted for and on addition of controls including children’s maths test score. With all controls, children in the lowest ‘ability’ group have 2.5 times the odds of negative self-concept compared to those in the highest group, and corresponding predicted probabilities of 15% compared to 7%.

2. Does the judgement by their class teacher of a child’s maths ability at age seven predict the child’s negative maths self-concept at 11?

Again, in the overall sample, teacher judgement of children’s maths ‘ability and attainment’ at seven predicts their maths self-concept at 11, accounting for ‘ability’ group, maths score, and other potential confounders. With all controls, children judged ‘below average’ have odds 3.5 times higher than those judged ‘above average’ of reporting not being good at maths at 11 – again, a substantive difference in predicted probabilities of 20% compared to 7%.

3. Do these relationships vary with a child’s early concurrent maths skill (as measured by maths cognitive test score at age seven)?

In the sample overall, the relationship between maths skill, as proxied by test score, and self-concept varies according to ‘ability’ group level, indicating that the impact of ‘ability’ group placement may differ for children with different current maths capability. However, the association of teacher judgements with later negative maths self-concept does not appear to vary with children’s maths skills.

4. Do these relationships vary by gender?

There are differences in relationships between ‘ability’ group and self-concept across girls and boys, particularly when analyses allow variation by maths test score. All high-group boys – regardless of score – have very low odds of reporting subsequently that they are not good at maths, while only high-scoring, high-group girls mirror this low probability. Low-scoring, high-group girls are more likely to have later negative maths self-concept. There is also some variation in the relationship between teacher judgements and self-concept for boys and girls of different concurrent skill levels. Girls judged ‘below average’ are more likely to have negative maths self-concept at 11, regardless of manifest maths skills at seven.

This suggests that different mechanisms and processes may mediate relationships between maths 'ability' group placement and maths self-concept for girls and for boys. Coupled with the apparently more unvarying relationship between negative teacher judgement and subsequent negative self-concept for girls, and the overall tendency – demonstrated through previous research and again in this sample – of boys more often to have positive maths self-concept than girls, it is feasible that girls and boys may be differentially sensitive to structural and social influences within the school environment on maths self-concept. Alongside this, the overall results for the whole sample support previous research indicating a stratifying effect of 'ability' grouping on self-concept and suggest a direct and lasting impact of teacher judgements, at the aggregate level. The subgroup analyses provide detail of the differential routes through which these factors may shape children's trajectories, beneath that aggregate.

### ***Differential effects of maths 'ability' group on the self-concept of girls and boys***

The findings of heterogeneous relationships by gender between maths in-class 'ability' group at seven and maths self-concept at 11 beg more questions than the MCS data can answer. Why do girls with relatively lower concurrent maths skills placed in the highest group have a higher probability of subsequent negative self-concept: an apparent transposition of the big-fish-little-pond effect not observed for sample boys? Why do sample boys, in contrast, appear to be impervious to contrast effects within their pond, and seem more straightforwardly to assimilate and absorb the label of their situation?

Previous research on 'ability' grouping tentatively provides the beginnings of some answers to these questions. Interviewing primary school children 'ability' grouped at different levels, Hallam et al. (2004) report experiences of higher placement that are not uniformly positive, describing 'pressure' among and negative social processes for some in the top group. In 1997, Boaler investigated top-set secondary school pupils, and describes an 'air of urgency' (p. 172) throughout lessons which consistently 'ignore[d] the individual needs of students' (p. 173). A number of girls in Boaler's study were left 'lost, confused and unhappy' (p. 176) by top-set pedagogy. Boaler cites research suggesting that girls tend to thrive in environments that are, 'non-confrontational and non-competitive' (p. 179), in contrast to those observed for her top-group pupils. Drawing also on work by Dweck, which suggests that 'tendencies toward unduly low expectations, challenge avoidance, ability attributions for failure, and debilitation under failure have been especially noted in girls' (p. 176), Boaler concludes that 'gender imbalance in the school mathematics system ... may be caused by certain features of the top set environment'. The possibility, then, is that early top-group placement has had a cumulative detrimental effect on the subsequent self-concept of those MCS girls whose skills were relatively less advanced at seven.

Carey et al.'s (2019) research into maths anxiety also supports the possibility of disadvantageous psychological effects for girls, with some female interviewees reporting a negative association between top maths 'ability' group and self-concept. One describes how 'my confidence just went straight down because I realised how clever everyone else was' (p. 45); another reports that 'I've always

been in the higher sets and there's always been people that are better' (p. 45). Congruent with findings from Boaler's (1997) study, girls in Carey et al.'s report relief on moving from the top maths 'ability' group to a lower placement: 'I'd feel like the teacher would kind of pressurise me ... rushing us ... the new teacher is nice, and she doesn't seem to rush me' (pp. 47–48).

The prospect raised by results here and by previous studies is therefore that as well as their overall stratifying effects, maths 'ability' groups have more complex implications for inequities by gender, with top group membership disadvantaging the self-concept of some sample girls – but not, seemingly, boys – leaving those girls who are (at the time of measurement) relatively less skilled, or developed, potentially more vulnerable to the negative effects of higher placement. Additionally, it is feasible that, given the established tendency of boys at the aggregate level to have more positive maths self-concept than girls (which is suggested again here by the low probability of negative self-concept among low-grouped but high scoring boys; [Figure 3](#)), and given corresponding stereotypes about gendered capabilities (Carey et al., 2019), only girls with higher concurrent skills are able cognitively to embrace and accept the notion of their own relative competence at maths conferred by high group placement. For girls whose skills have not yet progressed to the same stage, cognitive dissonance and insecurity might arise, leading to a lowered sense of self-competence.

### ***Teacher judgements and self-concept***

Turning to findings on teacher judgement, results indicate a relationship between early teacher ratings and children's later self-concept that is of a substantial magnitude. A key question, which cannot fully be addressed by the MCS data,<sup>7</sup> is whether the sample teachers' reported judgements of MCS children's maths skills represent a relative assessment of the child compared to their peers that is grounded or bears some accuracy, or whether, instead, it reflects tendencies to positive or negative perceptions on the part of the teacher.

Previous research has indicated that the judgements of MCS teachers are biased according to children's characteristics, and that boys who, at age seven, score equally to girls on the maths cognitive test are more likely to be judged 'above average' (Campbell, 2015). This provides evidence that these judgements are not simply reflective of the child within a concrete frame of reference, and supports the possibility that the rating of the child as 'above' or 'below' average reflects at least in part the teacher's own cognitive leanings. Moreover, given that attenuated models in the current paper control for children's maths skills – as proxied by the cognitive test – and for skills in other domains, as well as for background characteristics, this again suggests that patterns of ratings are at least to some extent situated at the level of the teacher: because variation in judgement remains after attenuation, and apparently similar children are judged differently.

Rubie-Davies (2007, 2010) shows a tendency of individual teachers to default to 'high' or 'low-expectation' thinking, and that 'high-expectation teachers spent more time providing a framework for students' learning, provided their students with more feedback, questioned their students using more higher-order questions, and managed their students' behaviour more positively' (p. 289). These details on the strategies of high-

expectations teachers may provide some explanation for the association found here between teacher judgements and children's later self-concept. If a teacher who tends to perceive and rate children more positively supports them with more a constructive and enabling classroom environment – and vice versa – this may have a long-run impact, including on self-concept.

If judgement style is inherent to the teacher to some extent, it is therefore worth concentrating resources and initiatives for change at this level, among those teachers with a tendency to view their pupils negatively. Findings here thus emphasise the need to take seriously the impact of teacher judgements on different aspects of children's experience, particularly in the context of inequalities in judgement by gender, of analyses in this paper suggesting a more pervasive association between unfavourable judgement and girls' self-concept, and given the wider context of under-attainment of girls in maths.

### ***Limitations and future research***

One limitation of the current research is the capacity of the maths cognitive test to measure children's skills. This is one test, taken at one time point, and subject to all the caveats regarding reliability and validity of any similar instrument (Harlen, 2007). It is possible that disparities and interactions conditional on test score level may to some extent be an artefact of test measurement error. But the question then remains: why would this play out differently for boys and girls? There is no obvious reason to think that girls placed in the highest maths 'ability' group, for example, would be more likely to have inaccurate test scores compared to boys placed at this level – and therefore interpretations of differences by gender and skill level are unlikely to be affected by this caveat.

Further limitations of the MCS data in answering some of the questions raised by findings here have already been mentioned. It is not possible to incorporate school composition into the current analyses, because of the lack of clustering of children within schools (the mean average is two) – though this may be addressed in future work when linked administrative data on school-make-up become available. In addition, as the data only exist for two time points – when children were aged seven, and 11, and as no reliable measure of self-concept is available at seven, it is not possible to track change, or, as discussed, specifically to examine mechanisms and mediators. Information on 'ability' groupings is collected at age 11, during wave five of the MCS, but, crucially, at a time point after the children report their self-concept – because the teacher survey once more follows fieldwork with families. Therefore, it is not possible validly to compare or interact associations between earlier and more recent grouping and maths self-concept.

Notwithstanding this, the magnitude and consistency of relationships indicated by this research illustrates a substantial potential 'snowballing' (Francis et al., 2020) of early maths in-class 'ability' grouping, and an enduring apparent effect of teachers' judgements, four years after their measurement (though the data do not allow detailed analyses of their interplay and dynamic interaction with one another). Future investigations will explore whether findings here are mirrored in alternative samples from different populations (which will address the limitation that research here is with one sample from one cohort of children), whether relationships of 'ability' group and teacher judgement with

maths self-concept continue to hold for the MCS children as they progress into secondary school, and whether there are implications for attainment and academic progress.

## Conclusions

Using a large, national sample of primary-aged children, this research set out to explore the relationships between early in-class ‘ability’ grouping for maths, early teacher judgements of children’s maths ability, and children’s later maths self-concept. It looked also at whether associations differ for girls and boys, as there are known disparities by gender in maths self-concept, and in related educational choices and careers, and there is therefore an imperative to understand factors that may be instrumental in these disparities. This is particularly important in the context of a ‘mathematics crisis’ in the UK, where overall capability among the population appears to be declining (Carey et al., 2019).

Analyses find that both ‘ability’ group and teacher judgement are strongly, independently related to later self-concept. The complex relationships between maths in-class ‘ability’ group and self-concept for girls, alongside the aggregate association of group with self-concept, once more invite acknowledgement by policymakers and practitioners and exploration of the use and impacts of ‘ability’-groupings among young children. In terms of teacher judgements, continued interrogation of the pedagogies and behaviours of low-expectation and high-expectation teachers may be fruitful, alongside further research into the reason that negative teacher judgement appears deleterious for the maths self-concept of girls regardless of skill level.

Both ‘ability’ group and teacher judgement are supported by this research as feasibly instrumental in forming primary children’s maths self-concept, in ways that vary by gender. Therefore both should be considered as sites for intervention which could boost maths progression and contribute to closing gender gaps.

## Notes

1. The majority of sample children were in year two at wave four (Hansen, Jones, Joshi, & Budge, 2010) and year six at wave five (Platt, 2014).
2. [https://cls.ucl.ac.uk/wp-content/uploads/2017/07/MCS5\\_MS\\_Child-Self-Completion-Questionnaire\\_CORE\\_ESNI\\_FINAL\\_PRINT.pdf](https://cls.ucl.ac.uk/wp-content/uploads/2017/07/MCS5_MS_Child-Self-Completion-Questionnaire_CORE_ESNI_FINAL_PRINT.pdf)
3. <https://cls.ucl.ac.uk/wp-content/uploads/2017/07/MCS4-Teacher-Technical-Report-v10.pdf>
4. Note that teachers also reported on other forms of grouping (‘setting’ and ‘streaming’) at age seven, and that there are co-occurrences between practices, with a strong correspondence between reported placement levels (e.g. children placed in the lowest in-class maths group are likely also to be in the lowest maths set – though some are NOT also set). In-class grouping for maths and for literacy are the most commonly occurring types of grouping at wave four, and also the most proximal, so this level of ‘ability’ grouping is the focus in this paper. See Campbell (2013) for further detail on different reported ‘ability’ groupings among the MCS children.
5. All specifications control for time lapse between the parent and child interviews at wave four and the subsequent teacher survey.
6. Note that though Key Stage One Scores are available for a subsample of MCS children, these scores are not included as controls, because they did not precede teacher judgements and ‘ability’-groupings: in the majority of cases, Key Stage One assessments took place after MCS wave four data collection, which fell during year two (Campbell, 2017; Hansen et al., 2010).

## Acknowledgements

I am grateful to The Centre for Longitudinal Studies, Institute of Education, for the use of these data, to the UK Data Archive and Economic and Social Data Service for making them available, and to all participants in the MCS. All analyses, interpretations and errors are my own. Many thanks to Ludovica Gambaro, Polina Obolenskaya and two anonymous reviewers for useful comments and feedback.

## Disclosure statement

No potential conflict of interest was reported by the author.

## Funding

This work was supported by a British Academy Postdoctoral Fellowship, under Grant [PF2 \180019].

## References

- Ansalone, G. (2003). Poverty, tracking, and the social construction of failure: International perspectives on tracking. *Journal of Children and Poverty*, 9(1), 3–20.
- Becker, H. (1963). *Outsiders*. New York: Free Press.
- Boaler, J. (1997). When even the winners are losers: Evaluating the experiences of top set students. *Journal of Curriculum Studies*, 29(2), 165–182.
- Boaler, J., Wiliam, D., & Brown, M. (2000). Students' experience of ability grouping - disaffection, polarisation and the construction of failure. *British Educational Research Journal*, 26(5), 631–648.
- Bradbury, A., & Roberts-Holmes, G. (2017). *Grouping in early years and key stage 1: "A necessary evil"?* London: UCL Institute of Education.
- Campbell, T. (2013). Stratified at seven: In-class ability grouping and the relative age effect. *British Educational Research Journal*, 40(5), 749–771.
- Campbell, T. (2015). Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment. *Journal of Social Policy*, 44(3), 517–547.
- Campbell, T. (2017). The relationship between stream placement and teachers' judgements of pupils: Evidence from the Millennium Cohort Study. *London Review of Education*, 15(3), 505–522.
- Carey, E., Devine, A., Hill, F., Dowker, A., McLellan, R., & Szucs, D. (2019). *Understanding mathematics anxiety: Investigating the experiences of UK primary and secondary school students*. Cambridge: University of Cambridge.
- Chui, M.-S. (2016). Effects of teacher assessment and cognitive ability on self-concepts: Longitudinal mechanisms for children from diverse backgrounds. *Saudi Journal of Engineering and Technology*, 1(4), 180–189.
- Codiroli McMaster, N. (2017). Who studies STEM subjects at A level and degree in England? An investigation into the intersections between students' family background, gender and ethnicity in determining choice. *British Educational Research Journal*, 43(3), 528–553.
- Connelly, R. (2013). *Interpreting test scores*. London: Centre for Longitudinal Studies.
- Connolly, P., Taylor, B., Francis, B., Archer, L., Hodgen, J., Mazenod, A., & Tereshchenko, A. (2020). The misallocation of students to academic sets in maths: A study of secondary schools in England. *British Educational Research Journal*, 45(4), 873–897.
- Francis, B., Archer, L., Hodgen, J., Pepper, D., Taylor, B., & Travers, M.-C. (2017a). Exploring the relative lack of impact of research on 'ability grouping' in England: A discourse analytic account. *Cambridge Journal of Education*, 47(1), 1–17.



- Francis, B., Connolly, P., Archer, L., Hodgen, J., Mazenod, A., Pepper, D., ... Tereshchenko, A. (2017). Attainment grouping as self-fulfilling prophesy? A mixed methods exploration of self confidence and set level among year 7 students. *International Journal of Educational Research*, 86, 96–108.
- Francis, B., Craig, N., Hodgen, J., Taylor, B., Tereshchenko, A., Connolly, P., ... Archer, L. (2020). The impact of tracking by attainment on pupil self-confidence over time: Demonstrating the accumulative impact of self-fulfilling prophecy. *British Journal of Sociology of Education*, 41(5), 626–642.
- Gripton, C. (2020). Children's lived experiences of 'ability' in the key stage one classroom: Life on the 'tricky table'. *Cambridge Journal of Education*, 50(5), 559–578.
- Hallam, S., Ireson, J., & Davies, J. (2004). Primary pupils' experiences of different types of grouping in school. *British Educational Research Journal*, 30(4), 515–533.
- Hallam, S., & Parsons, S. (2012). The incidence and make up of ability grouped sets in the UK primary school. *Research Papers in Education*, 28(4), 393–420.
- Hansen, K., & Henderson, M. (2019). Does academic self-concept drive academic achievement? *Oxford Review of Education*, 45(5), 657–672.
- Hansen, K., Jones, E., Joshi, H., & Budge, D. (2010). *Millennium Cohort Study fourth survey: A user's guide to initial findings*. London: Centre for Longitudinal Studies.
- Harlen, W. (2007). *The quality of learning: Assessment alternatives for primary education, primary review research survey 3/4*. Cambridge: University of Cambridge Faculty of Education.
- Heyder, A., Steinmayr, R., & Kessels, U. (2019). Do teachers' beliefs about math aptitude and brilliance explain gender differences in children's math ability self-concept? *Frontiers in Education*, 4(34), 1–11.
- Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford Review of Education*, 3(3), 343–358.
- Ireson, J., & Hallam, S. (2009). Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. *Learning and Instruction*, 19(3), 201–213.
- Johnston, O., Wildy, H., & Shand, J. (2019). A decade of teacher expectations research 2008–2018: Historical foundations, new developments, and future pathways. *Australian Journal of Education*, 63(1), 44–73.
- Jussim, L., & Harber, K. (2005). Teacher expectations and self-fulfilling prophecies: Knowns and unknowns, resolved and unresolved controversies. *Personality and Social Psychology Review*, 9(2), 131–155.
- Kutnick, P., Hodgkinson, S., Sebba, J., Humphreys, S., Galton, M., Steward, S., ... Baines, E. (2005). *Pupil grouping strategies and practices at key stage 2 and 3: Case studies of 24 schools in England*. London: Department for Education and Skills.
- Lazarides, R., & Lauermaann, F. (2019). Gendered paths into STEM-related and language-related careers: Girls' and boys' motivational beliefs and career plans in math and language arts. *Frontiers in Psychology*, 10, 1–17.
- Marks, R. (2014). The Dinosaur in the classroom: What we stand to lose through ability-grouping in the primary school. *FORUM*, 56(1), 45.
- Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96(1), 56–67.
- Marsh, H. W., Ludtke, O., Nagengast, B., Trautwein, U., Abduljabbar, A. S., Abdelfattah, F., & Jansen, M. (2015). Dimensional comparison theory: Paradoxical relations between self-beliefs and achievements in multiple domains. *Learning and Instruction*, 35, 16–32.
- Marsh, H. W., Pekrun, R., Murayama, K., Arens, K. A., Parker, P. D., Guo, J., & Dicke, T. (2018). An integrated model of academic self-concept development: Academic self-concept, grades, test scores, and tracking over six years. *Developmental Psychology*, 54(2), 263–280.
- McGuillicuddy, D., & Devine, D. (2020). 'You feel ashamed that you are not in the higher group'—Children's psychosocial response to ability grouping in primary school. *British Educational Research Journal*, 46(3), 553–573.

- Parsons, S., & Hallam, S. (2014). The impact of streaming on attainment at age seven: Evidence from the Millennium Cohort Study. *Oxford Review of Education*, 40(5), 567–589.
- Platt, L. (2014). *Millennium Cohort Study. Initial findings from the age 11 survey*. London: Centre for Longitudinal Studies.
- Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender and Society*, 26(2), 290–322.
- Rosenthal, R., & Jacobsen, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1), 16–20.
- Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology*, 77(2), 289–306.
- Rubie-Davies, C. M. (2010). "Teacher expectations and perceptions of student characteristics: Is there a relationship?". *British Journal of Educational Psychology*, 80(1), 121–135.
- Sukhandan, L., & Lee, B. (1998). *Streaming, setting and grouping by ability: A review of the literature*. Slough: National Foundation for Educational Research.
- Sullivan, A. (2013). Academic self-concept, gender and single-sex schooling. *British Educational Research Journal*, 35(2), 259–288.
- Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics*, 50(1), 49–62.
- Timmermans, A., Rubie-Davies, C. M., & Rjosk, C. (2018). Pygmalion's 50th anniversary: The state of the art in teacher expectation research. *Educational Research and Evaluation*, 24(3–5), 91–98.
- Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, 24(3–5), 124–179.