Leading education and social research
Institute of Education
University of London

Stereotyped at seven?

Biases in teacher judgements of pupils' ability and attainment

Tammy Campbell

CLS working paper 2013/8

September 2013

Centre for Longitudinal Studies
Following lives from birth and through the adult years
www.cls.ioe.ac.uk

CLS is an ESRC Resource Centre based at the Institute of Education, University of London

ESRC ECONOMIC & SOCIAL RESEARCH COUNCIL

Stereotyped at seven?
Biases in teacher judgements of pupils' ability and attainment

Tammy Campbell

September 2013

First published in 2013 by the
Centre for Longitudinal Studies
Institute of Education, University of London
20 Bedford Way
London WC1H 0AL
www.cls.ioe.ac.uk

ISBN 978-1-906929-64-0

The Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council (ESRC) Resource Centre based at the Institution of Education (IOE), University of London. It manages three internationally-renowned birth cohort studies: the 1958 National Child Development Study, the 1970 British Cohort Study and the Millennium Cohort Study. For more information, visit www.cls.ioe.ac.uk.

This document is available in alternative formats.
Please contact the Centre for Longitudinal Studies.
tel: +44 (0)20 7612 6875
email: info@cls.ioe.ac.uk

Contents

Acknowledgements

This paper presents analysis forming part of my PhD, which I am very grateful to the Economic and Social Research Council for funding. It uses data from the Millennium Cohort Study. I am grateful to the Centre for Longitudinal Studies, Institute of Education for the use of these data, and to the UK Data Archive and Economic and Social Data Service for making them available. However, these organisations bear no responsibility for the analysis or interpretation of these data.

Thanks to my supervisors, Lucinda Platt and Lorraine Dearden, for their continued critique and support. Many thanks also to Ellen Greaves for useful feedback on an earlier draft, to Meghan Rainsberry for asking the right questions, and to Nichola Shackleton and Dick Wiggins for constructive comments on initial analyses.

Summary

The Department for Education reports long-standing gaps in attainment among primary school pupils according to the following characteristics: family income level, gender, Special Educational Needs (SEN) status, ethnicity, and language(s) spoken[1]. This paper presents work in progress to investigate whether biases in teachers' assessments of pupils may contribute to creating or maintaining these attainment gaps among primary school children in England.

The research analyses data from a sample of more than 5,000 pupils and their teachers taking part in the national Millennium Cohort Study (MCS)[2].Teachers' assessments of the cohort members' reading and maths ability and attainment at age seven are compared to the children's independent performance in cognitive tests[3]. The research examines whether teachers' assessments differ according to each key characteristic (income / gender / SEN / ethnicity / language) for children with similar scores on the tests. The aim is to identify if and where there appear to be general biases in perceptions of each group. For example, the analysis examines whether teachers are more likely to judge boys or girls as better at reading, where both groups have scored equivalently on the reading test.

Main findings

In the MCS sample, the following groups are **less** likely to be judged by their teachers as 'above average' in **reading** than their equally scoring peers:

- pupils from low-income families
- boys
- pupils with any SEN diagnosis
- children who speak languages in addition to English
- Indian, Pakistani, Bangladeshi, Black Caribbean, and Black African pupils.

Correspondingly, nearly all of these pupil groups are more likely to be judged 'below average' at reading.

In the MCS sample, the following groups are **less** likely to be judged by their teachers as 'above average' in **maths** than their equally scoring peers:

- pupils from low-income families

[1]See, for example, http://www.education.gov.uk/rsgateway/DB/SFR/s001098/sfr30-2012.pdf; http://www.education.gov.uk/rsgateway/DB/SFR/s001104/sfr33-2012v2.pdf

[2] Data from Wave 4 (2008) is used. For more information on the study, see: http://cls.ioe.ac.uk/page.aspx?&sitesectionid=851&sitesectiontitle=Welcome+to+the+Millennium+Cohort+Study

[3] The cognitive tests are the British Ability Scales Word Reading Test and the Progress in Mathematics test. They were administered in children's homes by MCS survey interviewers. Teachers were entirely unaware of each child's test score when they provided their judgement of the child's 'ability and attainment.'

- girls
- pupils with any SEN diagnosis
- Black African and Black Caribbean pupils.

All but girls and Black African pupils are, correspondingly, more likely to be judged ‘below average’ at maths than their peers performing at the same level on the maths cognitive test.

The author then considers whether there is any link between local area and apparent biases in teacher assessments. It finds no change in teachers’ apparent biases when controlling for government office region, and biases are also evident in areas with greater socioeconomic and ethnic diversity.

The strongest indications that bias may be occurring nationally are for differences according to income level, gender and SEN status. In the MCS, pupils of non-White ethnicities and pupils speaking languages in addition to English are under-represented in some regions, meaning that the generalisability of these particular results to the national level cannot be determined.

Lastly, the author explores whether income level is in fact the primary driver of differences according to each of the other characteristics: gender, SEN, language, and ethnicity. It finds that controlling for family income level does not explain the majority of bias according to gender, SEN, language, or ethnicity. There seem to be independent biases in teacher perceptions of MCS pupils according to these characteristics, even where pupils have scored equivalently in cognitive tests and are from families with similar income levels.

This paper concludes that there is evidence to suggest stereotyping of pupils according to their group characteristics – in particular according to income level, gender and SEN status. Crucially, the paper does *not* conclude that there is anything unusual about teachers in their apparent tendency to stereotype pupils: stereotyping is a universal human process[4]. Instead, it recommends that renewed and increased credibility and importance be given to the growing body of evidence that biased judgements may be shaping pupil trajectories and affecting attainment. Research suggests that it is possible to tackle and alleviate stereotyping[5]. Therefore resources should be directed to enabling teachers (and other workers in education) to recognise, challenge and address the process – thereby potentially increasing parity in pupil attainment.

[4] See, for example, https://implicit.harvard.edu/implicit/demo/background/posttestinfo.html
[5] For example, http://www.wtamu.edu/webres/File/Journals/MCJ/earp.pdf

Introduction

Attainment gaps among primary school pupils

Department for Education (DfE) statistics describe key pupil-level characteristics which underpin gaps in average attainment throughout primary school - from the first year (where Foundation Stage Profile [FSP] scores are awarded) to the last (where Key Stage Two [KS2] examinations are completed). In English state-maintained schools, the main characteristics according to which gaps are reported are:

- income-level (whether a pupil is in receipt of free school meals [FSM]);
- pupil gender;
- whether a pupil has a special educational needs diagnosis (SEN);
- pupil ethnicity;
- whether a pupil speaks languages in addition to English (EAL).

At the foundation stage, pupils of White and Indian ethnicities are reported as achieving at a higher level than the average, and Pakistani, Bangladeshi, Black Caribbean and Black African pupils below the average. Girls score at a higher level than boys, and children classified as speaking English only at a higher level than those who speak languages in addition to English. Pupils recorded as in receipt of FSM and pupils with any recognised SEN diagnosis also attain at relatively lower levels.

At Key Stage One (KS1), when pupils are aged seven, and at KS2, achievement gaps mirror those reported at the foundation stage – with the exception of a systematic gender distinction, where boys score higher than girls in maths, and girls higher than boys in English (DfE 2012a; 2011; 2012b).

Bias in teacher assessment

Research indicates that these attainment differentials may be due, in part, to teacher over- and under-assessment of pupils on the basis of their characteristics and group memberships. While FSP and KS1 results are purely based on in-school assessment by each class teacher, KS2 results combine scores from externally marked examinations with teacher judgements.[6] In a 2009 paper, Burgess and Greaves exploit this distinction and, using comprehensive English national data, compare the externally-marked and teacher-assessed components of the KS2 levels awarded to pupils. They find indications of systematic biases in teacher assessment which are largely in line with DfE-reported attainment gaps: *over*-assessment of Indian pupils, and *under*-assessment of Black Caribbean and Black African pupils, of pupils in

[6] http://www.education.gov.uk/schools/teachingandlearning/curriculum/a0068102/early-years-foundation-stage-eyfs;
http://www.education.gov.uk/schools/teachingandlearning/assessment/keystage2/a00203532/about-key-stage-2-assessment

receipt of FSM, and of pupils with SEN – suggesting that teacher bias serves to inflate and deflate the scores allocated to each pupil.
Analysing the English sub-sample of the Millennium Cohort Study (MCS), Hansen and Jones (2011) indicate that teachers may also be biased in their assessments of pupils at the beginning of primary school. They compare children's (age five) FSP scores to self-completed cognitive tests taken during the same academic year, outside of school, and find greater disparities according to gender in the teacher-assessed FSP measure than in the child-completed tests. Teacher assessments pronouncedly favour girls to an extent incongruent with a lesser gender differentiation in children's test performance, indicating that the variations by pupil group at the foundation stage may – like those at KS2 – be due in part to bias.

Evidence suggesting that there are biases in teacher assessments at primary level has built for over a decade (e.g. Reaves et al, 2001; Thomas et al 1998), along with an accumulation of research that indicates similar tendencies in secondary schools. For example, Strand (2012) finds that Black Caribbean pupils are significantly less likely to be entered into higher tiers for national tests at age 14 than White British counterparts with equivalent prior attainment and family characteristics.

Policy responses to attainment disparities

Despite this mounting evidence for a systematic, inequitable penalisation of certain pupil groups, explicit recognition and exploration of the possibility at the policy-level has been limited. Socio-economic status (SES), often according to the proxy of FSM, has frequently been cited as an explanatory factor which drives attainment gaps, but proposed mechanisms through which SES variation plays out in differentiated pupil attainment have tended to focus at the family-level: for example, emphasising the (evidenced) contribution of variation in home learning environment, and in pupil and parent attitudes (DFES, 2005; DCSF, 2008). The potential for explanation at the teacher-level has remained largely unacknowledged.

Though differences in pupil attainment are clearly delineated by characteristic in DfE (and in the preceding DCSF's) statistical reporting, and though a number of government-commissioned research projects have attempted to disentangle the reasons for (for example) variation according to pupil ethnicity (Maylor et al, 2009; Strand et al, 2010; Wilkin et al, 2010), policy has essentially ignored the prospect raised through these very studies that teacher perceptions and behaviours may play a part in creating differences. This is despite explicit suggestions to this effect: Maylor et al's (2009) evaluation for the DCSF of the Black Children's Achievement Programme concludes that:

> Institutional factors / processes including negative teacher attitudes / expectations and stereotypical thinking about the ability of Black children serve to undermine teacher ability to raise Black children's attainment at an individual and group level…and reports 'negative attitudes and prejudice among school staff' (p 2).

Similarly, Strand et al's (2010) investigation into *Drivers and Challenges in Raising the Achievement of Pupils from Bangladeshi, Somali and Turkish Backgrounds* concludes that:

> Racism and structural inequalities may be important influences on the attainment of many Bangladeshi and Somali students (p 18).

The procurement of these studies under the previous Labour government may have indicated the beginnings of an interest in considering and exploring explanations for variation in pupil attainment that incorporate teacher-level bias. However, the current administration's focus has returned to family-level poverty as the fundamental driver of inequalities:

> For far too long we have tolerated the moral outrage of an accepted correlation between wealth and achievement at school…Children on free school meals do significantly worse than their peers at every stage of their education (DfE, Schools White Paper, 2010a, p 4).

> At the heart of our Coalition's Programme for Government is a commitment to spend more money on the education of our poorest children (DfE White Paper Equalities Impact Assessment, 2010b, p 8).

The Coalition Government has abolished Contextual Value Added (CVA) scores[7] with an implied nod towards the importance of teacher perceptions, on the basis that:

> [CVA scores had] the effect of expecting different levels of progress from different groups of pupils on the basis of their ethnic background, or family circumstances, which we think is wrong in principle (DfE, 2010a, p 68)

However, this simultaneously marginalises both the corresponding longstanding (quantified and documented) differences in attainment according to a greater range of characteristics than income-level alone and the recent and historical evidence for processes and explanations related to various of these factors which lie at the teacher-level. The White Paper's Equalities Impact Assessment (DfE, 2010b) indicates, for example, that SEN and ethnicity are descriptors – but not sources – of between-pupil variation. The Pupil Premium, which channels funds to less wealthy pupils, is presented as a solution which, by targeting poverty, will alleviate many of the disparities according to ethnicity and SEN:

> As many deprived [children] also have Special Educational Needs or are members of underachieving ethnic groups…significant numbers of pupils from these groups will also benefit from the extra resources and tailored support the Pupil Premium will provide (p 9).

---

[7] CVA scores were based on an iterative, fluid statistical model which predicted pupil progress according to a number of factors, including pupil characteristics such as income-level, gender, SEN, ethnicity and EAL (DCSF, 2007).

Biased assessments through stereotyping

Parallel to – but seemingly set aside by – the (present) policy process, research has, as discussed, continued to provide evidence for the existence of biases in teacher assessment which may account significantly for some of the variation in recorded attainment. Research has also developed hypotheses for the means by which bias in teacher perceptions may emerge, suggesting stereotyping of pupils by teachers as a mechanism. Burgess and Greaves, for example, put forward a stereotype model to account for their finding of biases at KS2, and contend that 'the teacher combines information derived from observing…the specific, individual pupil…with the prototype for that pupil's group' (p 19).

There are a number of theories of what stereotypes *are*, and of behaviours associated with their presence. In social psychology, many are grounded in the premises that stereotypes comprise invariant, homogenous, evaluative judgements of a given group (e.g. social, gender or ethnic group), and that stereotypes enable judgements of a given group to be made quickly and with cognitive ease (see e.g. Hilton and von Hipple, 1996; McGarty et al, 2002.) By stereotyping, therefore, teacher judgements of pupils can be made quickly and with cognitive efficiency (though with compromised *accuracy*) based, in part, on a preconceived 'template' of the ability and attainment of low-income pupils, girls, boys, SEN pupils, and so on. No assumption that stereotyping takes place on a conscious or deliberate level (that teachers are being intentionally prejudiced) is necessary: the process's efficiency is engendered by its automaticity.

Several theorists argue furthermore that stereotypes must be held at the group or institutional level: '…stereotypes should be formed in line with the accepted views or norms of social groups that the perceiver belongs to' (McGarty et al, 2002, p 2). The possibility, therefore, is that teachers, as a profession, use normalised templates of pupil attainment, which are premised on pupil characteristics, and which inform assessments of each child.

Building upon previous research findings

Does the evidence to date unequivocally support the theory that pupils are being stereotyped by their teachers, and that it is this process which results in assessment bias and which contributes to differentiated attainment levels? Though they propose and support a stereotype model, Burgess and Greaves (2009) also acknowledge and outline a possible competing explanation for their findings. Because their analysis uses comparators from within the same overall system (the teacher who assesses the pupil also teaches them for the externally-marked test), there is a danger of relationships within this system. For example, they suggest, the notable difference between the teacher-assessed and externally-marked elements of SEN pupils' results, in particular, may be due to: '…an extreme form of "teaching to the test" for pupils with SEN…the teacher's more in-depth knowledge of the student's ability may result in a lower [teacher assessment]' (p 12). Teachers might explicitly train and focus on certain pupils, whom they see as less able, so that they learn to attain desirable KS2 levels – but, as a result, these levels may not reflect the teacher's day-to-day perception of the pupils' ability.

Hansen and Jones' (2011) analysis partially circumvents this issue by utilising measures of pupil 'ability' which are not explicitly related to their schooling, or directly influenced or reported by their teacher, and which are not assessed within the educational context. They use the cognitive tests independently administered in children's homes as part of the MCS as a comparator to school-based, teacher-assessed FSP scores, arguably providing an enhanced indication that teacher judgements are biased away from manifest pupil performance.

However, while this study strengthens the evidence that recorded teacher assessments are inaccurately skewed, a danger remains that the FSP scores used do not in fact comprise direct portrayals of the mental representation – the potentially stereotype-based 'evaluative judgement' – that each assessing teacher holds of their (groups of) pupils. Because schools themselves, at the institutional level, are judged by the attainment of their pupils, and because teachers' own performance is assessed according to the performance of their pupils (Bradbury 2011a), there is some danger that FSP scores serve not simply to describe the perceived attainment or progress of each individual child, but to inform other purposes: 'when a measure becomes a target, it ceases to be a good measure.'[8]

A recent report by Ofqual (2012) noted, for example, a tendency within teacher assessment to manipulate 'marks so that candidates [are] placed within certain perceived grade boundaries' (p 82), while one response to a 2009 Ofsted consultation on school inspections stated that: 'Schools can manipulate their CVA scores[9] in ways that Ofsted would be unlikely to support.'[10] Bradbury (2011b) describes findings from case studies where 'assessment results may be influenced by pressure from external advisors, who only recognise certain patterns of results as intelligible' (p 655). Recorded FSP scores may, therefore, provide an inaccurate representation of teacher perceptions of a given individual or group, due to their complicity with a system where the attainment levels awarded to pupils have implications far beyond measuring and assessing each child's ability, progress or performance.

In order, therefore, more unambiguously and explicitly to investigate whether it is stereotyping of pupils by teachers that underpins biased assessment according to pupil characteristics, the analysis presented in this paper uses a measure of teacher judgement which is not part of or required by the education and assessment system,

[8] 'Goodhart's Law', as paraphrased by Strathern, (1997).

[9] As discussed, CVA scores were abolished by the coalition government in 2010 (see DfE, 2010a), but were used in all state schools in the years previous to this (including those when the MCS data was collected). Schools were assessed on the basis of their achieved CVA scores, therefore providing an incentive for differential manipulations of these scores according to the pupil characteristics known to be accounted for in the CVA model (these included ethnicity, gender, EAL, FSM, and SEN [DCSF, 2007]).

[10] http://ofstednews.ofsted.gov.uk/article/346

which is removed from its context, and which will not inform evaluations of performance of a teacher or their school. Confidential MCS survey responses by teachers to questions about their pupils' 'ability and attainment' (at age seven) provide a proposed proxy for the mental representations held by each teacher. These survey responses should lack the agenda inherent to formal in-school scorings. In addition, like Hansen and Jones' analysis, the current research uses independent MCS-administered cognitive test scores (also collected at age seven) as indicators of each child's performance.

Furthermore, specifically to address the apparent tendency in educational policy-making to attribute attainment gaps primarily to differences in SES (specifically, income-level), analysis goes on to investigate whether income-level accounts for any apparent stereotyping according to each of the other characteristics which are reported as delineating pupil achievement (gender, SEN, ethnicity, EAL). Previous studies have suggested that the assumption of income-level as primary driver may not hold. For example, Strand's (2011) paper, *The limits of social class in explaining ethnic gaps in educational attainment,* found that differences in early secondary achievement between White British and Black Caribbean pupils could not be explained by variation in income-level, nor by a plethora of additional contextual factors. This paper will, therefore, explore whether controlling for income-level explains any bias according to the other key characteristics of the MCS seven-year-olds.

Research questions

The following overriding research questions will therefore be addressed:

- At age seven, are there biases in survey-reported teacher judgements of pupils corresponding to each of the key pupil characteristics underpinning recorded primary-age attainment gaps (family income-level, gender, SEN, ethnicity, EAL)?

- Does controlling for income-level explain any biases according to gender, SEN, ethnicity, or EAL?

The data

The Millennium Cohort Study included 11,695 English children at its first sweep in 2001, and four additional waves have taken place to date, in 2004, 2006, 2008, and 2012. Analysis here uses data from wave four, when the pupils were seven years old, and in year two at primary school. Analysis is restricted to children in England, in order to allow comparison with and interpretation in the context of DfE statistics on pupil attainment. Twins and triplets are removed from analysis, because teacher bias and stereotyping may follow a different process for these pupils. The sample is further restricted to include only children whose parents report *not* paying school fees, again, in order to ensure comparability with the DfE statistics for state school pupils in England.

Responses to the survey of teachers at wave four which provides much of the data used here were received for only a subsample of pupils. This further reduces the sample size. Figure 1 compares numbers in the original wave one sample to the wave four sub-sample with all necessary key data which is used in this paper. (There are some – usually minor – variations in sample sizes according to individual question non-response; exact numbers are reported throughout analysis.)

Figure 1: Sample sizes at each stage of the MCS

Comparability of samples

Clearly, the sub-sample used in this paper is substantially smaller than the original, representative English sample of babies born shortly after the Millennium. Because the survey weights developed for use with wave four MCS data are designed to correct for divergences from the population at the pupil rather than school or teacher-level, they cannot be used accurately to render data from the teacher survey representative of all teachers in England. Unweighted data is therefore used throughout this paper, and Table 1, below, roughly compares the sample of pupils who have teacher survey and cognitive test data to the schools population in England in 2008-09 as reported in (then) Department of Children, Schools and Families statistics (2009a).

Table 1: Pupil characteristics in the English MCS wave four (2008) teacher sample and in the English school population in 2008-09

Characteristic	**Measure / definition** **Proportions in 2008 MCS teacher sample (unweighted)**	**Measure / definition** **Proportions in state school population according to Department for Children, Schools and Families statistics for pupils in 2008-09**
Gender	Parent-report in survey	Statistics for pupils who were 7 in January 2009
	50.3% male	**51.2%** male
Ethnicity	Parent-report in survey / derived variable	Statistics for all state primary pupils
	80.8% White, 3.0% Indian, 7.3% Pakistani / Bangladeshi, 3.9% Black,1.9% ‘Other,’ 3.4% Mixed ethnicity	**79.2%** White, 2.5 % Indian, 5.5% Pakistani / Bangladeshi, 4.9 % Black, 3.8% ‘Other,’ 4.1% Mixed ethnicity
English as an additional language	Parent report in survey: Response to question on 'language spoken in household”	“First language is known or believed to be English” - statistic for all state primary pupils
	86.2% “English only”	**84.6%** English first language
Diagnosed / recognised with special educational needs (SEN)	Teacher report in survey: Response to question “Has this child EVER been recognised as having SEN?”	Pupils in year two in 2008-09
	23.2% “yes”	**21.8%** with any SEN recorded

No stark incongruities emerge according to these comparisons, but there remains a danger that, though the pupil-level sample appears similar to the population according to these characteristics, the teacher-level sample may be skewed compared to the population of teachers (and, of course, the sample and the population may differ according to unobserved characteristics). Unfortunately, DfE statistics on the schools workforce in the years around MCS wave four surveying do not allow meaningful comparisons with the sample – so any relationships found in the current analysis can be attributed with certainty only to the teachers and pupils included here. Findings from this paper can, however, validly be interpreted within the context of previous research (some of which is nationally representative - e.g. Burgess and Greaves, 2009) and used to theory-build and to explore whether the hypothesis that stereotyping by teachers takes place is supported – without necessarily assuming an exact representation of the teacher and schools population.

Methodology: Cognitive test scores and teacher judgements

MCS wave four teacher-reported judgements of whether a pupil is ‘well above average / above average / average / below average / well below average’ at reading and maths, respectively, are used. These evaluations are in response to a survey question asking the teacher to ‘rate [the given] aspect of the study child’s ability and attainment [reading / maths]…in relation to all children of this age…’[11]

Responses are recoded into binary variables representing a rating of ‘above’ or ‘below’ average, which indicate whether each child is judged as relatively more or less able, compared to their peers. Responses of *well above average* and *above average* are combined to form an ‘above average’ category, where all else is categorised ‘not above average;’ similarly responses of *well below average* and *below average* are combined to one ‘below average’ category. Four outcome variables are thereby created:

- teacher judgement of reading ‘above average’ / not;
- teacher judgement of reading ‘below average’ / not;
- teacher judgement of maths ‘above average’ / not;
- teacher judgement of maths ‘below average’ / not.

In addition, the following measures of each of the pupil characteristics identified by DfE statistics as underpinning attainment variation are used (all are taken at wave four):

- a derived variable from parent-reported data which indicates whether the family’s income is above / below an OEDC 60% of median UK income poverty indicator;
- parent-reported pupil gender;
- teacher report of any recognised SEN (yes / no);
- a derived variable from parent report denoting pupil ethnic group (White / Black African / Black Caribbean / Bangladeshi / Pakistani / Indian);
- a derived variable from parent-reported information on language(s) spoken in the pupil’s household (coded to represent English only / additional languages).

Further details on the origin of each variable and of any recoding can be found at Annex A. Variables were selected and coded with the intention of mirroring as closely as possible the categorisations used in DfE reporting.

Only sub-sets of breakdowns by ethnicity are reported in this paper, in order to aid meaningful interpretation and comparison with DfE statistics. The census-based eight-category ethnicity categorisation[12] is used throughout analysis, and includes

[11] See http://www.esds.ac.uk/doc/6848/mrdoc/pdf/mcs4_teacher_england.pdf for full survey documentation.

[12] This groups pupils according to the following categories: White, Mixed, Indian, Pakistani, Bangladeshi, Black Caribbean, Black African, Other Ethnic Group.

‘other’ and ‘mixed’ classifications – but results for these groups are not reported. Descriptive statistics according to ethnicity may therefore not sum to 100%, while in modelling, noted sample sizes are for the whole sample with ethnicity data – as all are included in analysis – although only results for selected groups are reported.

Table 2 shows the percentage of pupils with each characteristic who are evaluated as relatively more or less able than their peers, according to the definitions described above. It indicates a lower chance of being evaluated as ‘above average’ at reading for boys, pupils speaking languages in addition to English, pupils with SEN, lower-income pupils, and pupils of all ethnicities except White and Indian. The same pattern holds for judgements of maths ability, save for a reversal according to gender, with boys more highly rated here.

Table 2: Percentage of pupils with each characteristic judged at each level by their teacher

	Percentage judged ‘above average’ at reading	**Percentage judged ‘below average’ at reading**	**Percentage judged ‘above average’ at maths**	**Percentage judged ‘below average’ at maths**
Whole sample (5161 / 5146)	45.3	22.2	39.8	20.9
Boys (2594 / 2589)	40.3	27.1	42.1	21.9
Girls (2567 / 2557)	50.4	17.3	37.4	19.9
Speaks English only (4449 / 4435)	46.7	21.5	41.1	25.6
Speaks additional languages (712 / 711)	36.8	26.8	31.6	20.2
No SEN diagnosis (3926 / 3912)	55.9	9.2	48.6	9.5

	Percentage judged 'above average' at reading	Percentage judged 'below average' at reading	Percentage judged 'above average' at maths	Percentage judged 'below average' at maths
Any SEN diagnosis (1183 / 1182)	11.2	64.9	11.2	55.7
Above 60% median income (3693 / 3682)	55.4	16.3	45.7	16.1
Below 60% median income (1464 / 1460)	27.5	37.3	24.6	33.1
White (4131 / 4115)	47.1	21.2	41.4	19.8
Black African (116 / 116)	35.3	31.9	25.9	31.0
Black Caribbean (68 / 68)	32.4	35.3	22.1	41.2
Bangladeshi (88 / 88)	37.5	33.0	33.0	26.1
Pakistani (285 / 285)	29.5	29.5	23.9	31.2
Indian (155 / 155)	47.7	18.1	44.5	14.8

Cognitive test scores at age seven

During wave four of the MCS, when children were aged seven, they completed a number of cognitive tests, in their homes, during a visit from a survey administrator.

They included the British Ability Scale Word Reading test, and a shortened version of the Progress in Mathematics test.

The Word Reading test is designed to assess children's English reading ability (see http://www.gl-assessment.co.uk/products/bas3). The ability score (a scaled but not otherwise standardised score) is used in analysis.

The Progress in Mathematics test is designed to measure pupils' mathematical ability across use of numbers, shapes, and skill in data handling. It is intended to provide an indication of performance in maths at the given developmental stage (see http://www.gl-assessment.co.uk/products/progress-maths). The shortened version used in the survey entailed routing to sections of varying difficulty levels, and Rasch scaling was used to convert the raw scores to a count score equivalent to that which would be attained were the full test completed (see Hansen, 2012). It is this scaled score which is used in analysis here.

Performances on these cognitive tests provide points of comparison to the teacher assessments of pupil reading and maths 'ability and attainment.' Pupil completion of the cognitive tests preceded teacher completion of their survey, as permission for the latter was gained from children's families during the MCS interviewer's visit.[13] Comparability of the two measures necessitates assumptions that: a) the lag between pupil test completion and teacher survey completion does not vary systematically across the pupil characteristics of interest; b) children delineated by each of the characteristics of interest develop at equivalent rates in their reading and maths ability and performance, at age seven (so that any apparent bias in teacher assessments cannot be attributed to slower progress during the time lag from pupil survey to teacher survey in some groups). The second of these assumptions cannot explicitly be tested using the MCS data, so remains a supposition (though as the modal time lag was short, at two months, it seems reasonable); the first is supported by additional analysis available in Annex B.

BAS Word Reading scores

Table 3, below, shows the distribution of Word Reading scores across the sample of pupils who took the test and who also have responses to the teacher-completed question on reading ability, according to each characteristic of interest. It also shows the proportion of pupils in each group who performed at or below the 20th, 50th, and 80th percentile values for this whole sample. Figures 2 and 3 then illustrate the proportions performing at or below this 20th percentile value and above the 80th percentile value in order to distinguish differences in percentages of 'low' and 'high' scorers according to each pupil group.

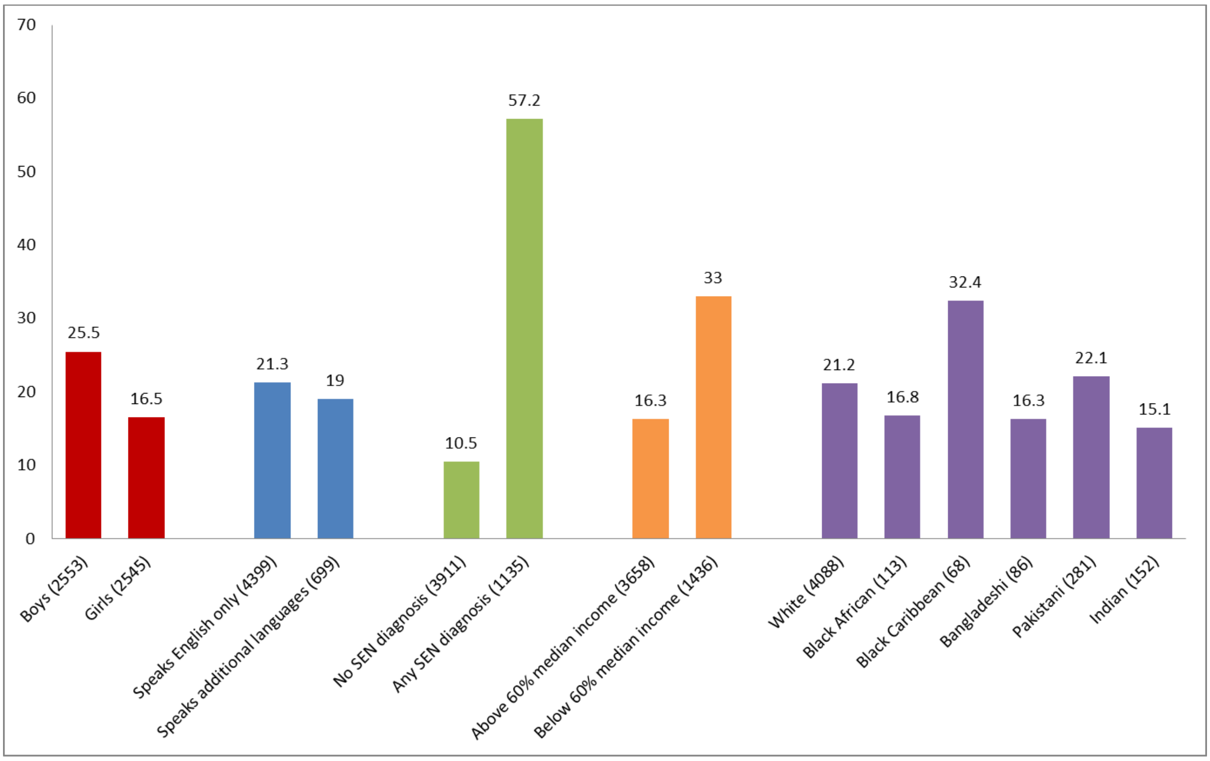
[13] The mean average time lag between cognitive test and teacher survey is 3.8 months, the median 3 months, and the mode 2 months.

Table 3: Distribution of Word Reading scores among pupils with each characteristic

	Mean	Median	20th percentile	50th percentile	80th percentile	Percentage of cases scoring at or below 20th percentile	Percentage of cases scoring at or below mean	Percentage of cases scoring at or below 80th percentile
Whole sample with teacher reading judgement (5098)	**109.5**	**111.0**	**86.0**	**111.0**	**134.0**			
Boys (2553)	106.9	108.0	81.0	108.0	134.0	25.5	55.1	80.2
Girls (2545)	112.1	114.0	91.0	114.0	134.0	16.5	47.1	81.1
Speaks English only (4399)	109.5	111.0	85.0	111.0	134.0	21.3	51.4	81.2
Speaks additional languages (699)	112.2	113.0	87.0	113.0	138.0	19.0	49.2	77.3
No SEN diagnosis (3911)	117.2	117.0	96.0	117.0	138.0	10.5	41.6	76.7
Any SEN diagnosis (1135)	83.2	82.0	60.0	82.0	106.0	57.2	83.5	94.3
Above 60% median income (3658)	113.4	114.0	91.0	114.0	136.0	16.3	46.3	78.4
Below 60% median income (1436)	99.7	100.0	73.0	100.0	127.0	33.0	63.3	86.4

	Mean	Median	20th percentile	50th percentile	80th percentile	Percentage of cases scoring at or below 20th percentile	Percentage of cases scoring at or below mean	Percentage of cases scoring at or below 80th percentile
White (4088)	108.9	111.0	85.0	111.0	134.0	21.2	51.6	81.3
Black African (113)	116.08	115.0	88.0	115.0	141.4	16.8	45.1	76.1
Black Caribbean (68)	103.7	105.5	76.0	105.5	134.0	32.4	60.3	82.4
Bangladeshi (86)	113.9	107.5	90.0	107.5	14.8	16.3	54.7	70.9
Pakistani (281)	109.0	111.0	85.0	111.0	132.0	22.1	50.7	81.5
Indian (152)	118.0	117.0	91.0	117.0	143.0	15.1	42.8	73.7

Figure 2: Percentage of pupils with each characteristic scoring at or below 86 (20th percentile score for whole sample) on the BAS Word Reading test ('Low scorers')*



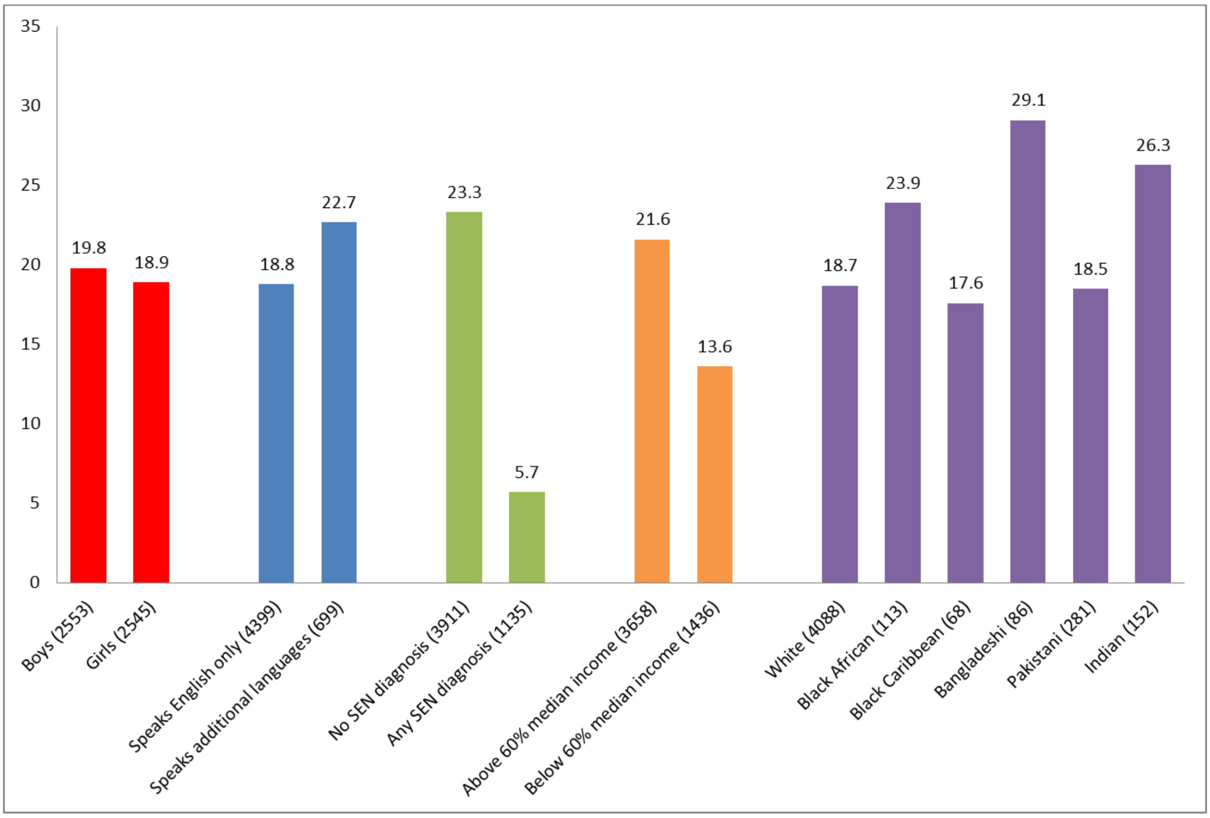
*Individual N in brackets after each characteristic

Figure 2 shows that more boys than girls scored at the bottom end of the distribution on the Word Reading test; that pupils with SEN were far more likely than pupils without SEN to score at this level; and that lower-income and Black Caribbean pupils are also more likely to be low scorers. Language spoken in the home appears to bear little relationship to low Word Reading score. Black African, Bangladeshi and Indian pupils are proportionally less likely to score at the bottom end on the Word Reading test.

Figure 3 shows no great gender difference in proportions of pupils scoring at the top end of the test. Pupils speaking languages in addition to English appear more likely to score highly, as well as Black African, Bangladeshi, and Indian pupils. Pupils with any SEN diagnosis are unlikely to attain a relatively high score – though 5.7% of the 1135 with recorded SEN do score highly – and low-income pupils are also less likely.

Though measured on different scales and not, therefore, directly comparable, these descriptive statistics begin to indicate incongruities between children's cognitive test scores and judgements by their teachers. Pupils speaking languages in addition to English appear more likely to score relatively well on the BAS Word Reading test – but are less likely than pupils speaking only English to be rated highly at reading by their teacher. Similarly, Black African and Bangladeshi pupils score relatively highly on the Word Reading test – but are less likely to be judged 'above average' and more likely to be judged 'below average' by their teacher.

Figure 3: Percentage of pupils with each characteristic scoring above 134 (80th percentile score for whole sample) on the BAS Word Reading test ('High scorers')*



*Individual N in brackets after each characteristic

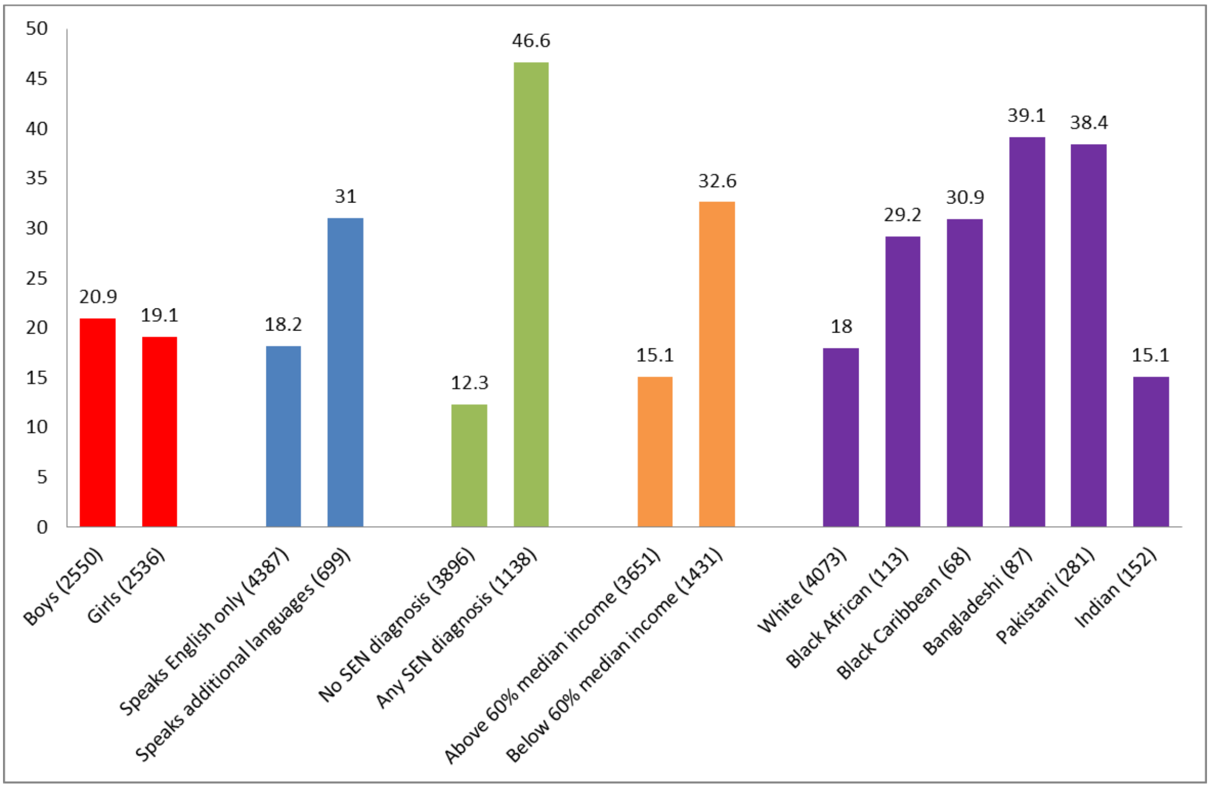
Progress in Mathematics scores

Table 4 shows the distribution of Progress in Mathematics scores across the whole sample of pupils with teacher judgements on maths ability, according to each characteristic of interest. Again, it shows the proportion of pupils in each group who performed at or below the 20th, 50th, and 80th percentile values for the whole sample. Figures 4 and 5 then illustrate the proportions performing at or below the 20th percentile and above the 80th percentile.

Table 4: Distribution of Progress in Maths scores among pupils with each characteristic

	Mean	Median	20th percentile	50th percentile	80th percentile	Percentage of cases scoring at or below 20th percentile	Percentage of cases scoring at or below mean	Percentage of cases scoring at or below 80th percentile
Whole sample with teacher maths judgement (5086)	**18.46**	**20**	**12.8**	**20**	**24**			
Boys (2550)	18.46	20	12	20	24	20.9	56.3	83.4
Girls (2536)	18.46	20	14	20	23	19.1	58.0	89.7
Speaks English only (4387)	18.69	20	14	20	24	18.2	56.2	87.6
Speaks additional languages (699)	16.97	17	10	17	23	31.0	63.1	90.7
No SEN diagnosis (3896)	19.65	20	15	20	24	12.3	50.4	85.9
Any SEN diagnosis (1138)	14.36	14	9	14	20.2	46.6	80.1	95.0
Above 60% median income (3651)	19.27	20	14	20	24	15.1	52.3	86.2
Below 60% median income (1431)	16.36	17	10	17	23	32.6	69.5	92.7
White (4073)	18.73	20	14	20	24	18.0	55.9	87.9
Black African (113)	17.34	18	10	18	24	29.2	59.3	88.5
Black Caribbean (68)	17.25	17.5	11.6	17.5	23.2	30.9	64.7	88.2
Bangladeshi (87)	15.54	16	10	16	21	39.1	71.3	96.6
Pakistani (281)	15.55	16	9.4	16	21	38.4	71.9	95.0
Indian (152)	19.88	21	14	21	26	15.1	46.7	75.7

Figure 4: Percentage of pupils with each characteristic scoring at or below 12.8 (20th percentile score for whole sample) on the Progress in Mathematics assessment ('Low scorers')*

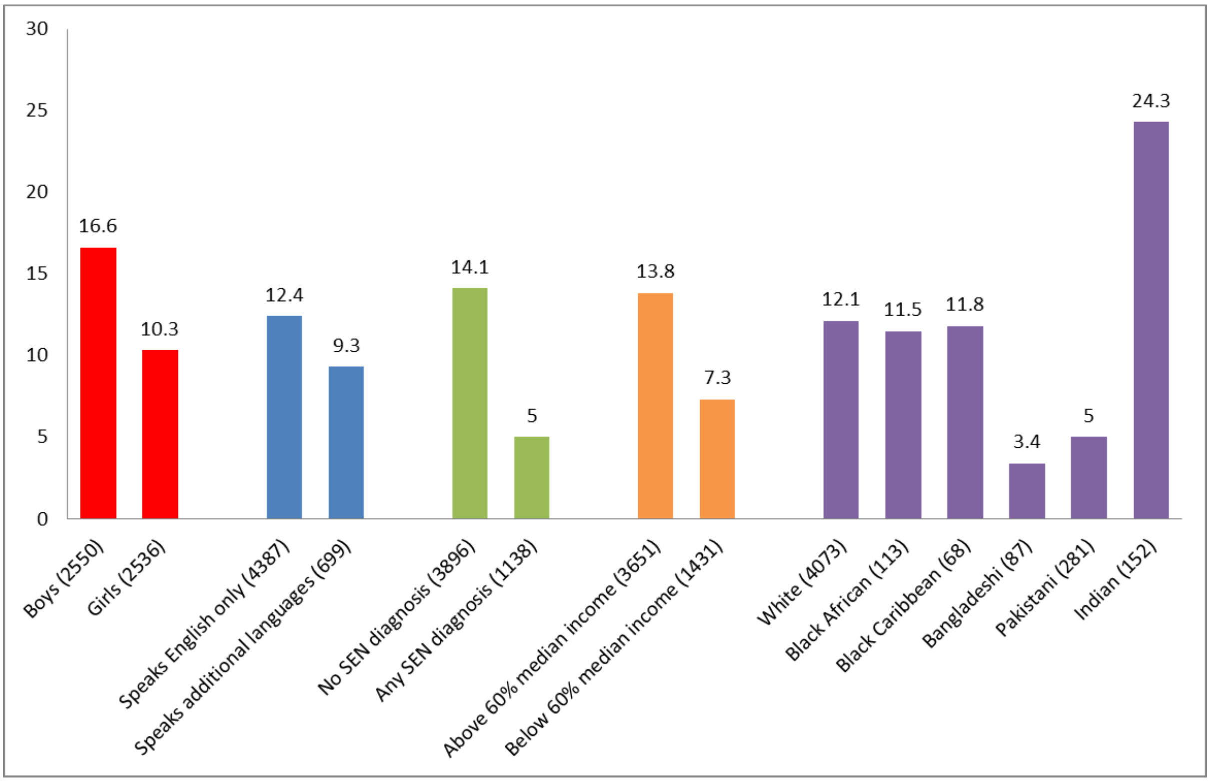


*Individual N in brackets after each characteristic

As with Word Reading scores, Figure 4 shows that sample pupils with SEN and less wealthy pupils are more likely to score at the bottom end of the Progress in Mathematics test. In contrast to Word Reading, however, boys are slightly more likely than girls to attain a low score, and pupils speaking languages in addition to English are also disproportionately represented here. Pupils of all reported ethnicities except for White and Indian are relatively more likely to attain a lower score on this test.

As well as being overrepresented at the low end, Figure 5 shows that boys are more likely to score at the high end of the Progress in Mathematics assessment – along with pupils who speak only English and, to a notable extent, Indian pupils. Less wealthy pupils and pupils with SEN are unlikely to perform at the top level here – though, again, there are some pupils with SEN scoring highly (5% of 1138). Bangladeshi and Pakistani pupils are comparatively unlikely to score at this top level. Black Caribbean and Black African pupils are also slightly underrepresented here.

Figure 5: Percentage of pupils with each characteristic scoring above 24 (80th percentile score for whole sample) on the Progress in Mathematics assessment ('High scorers')*



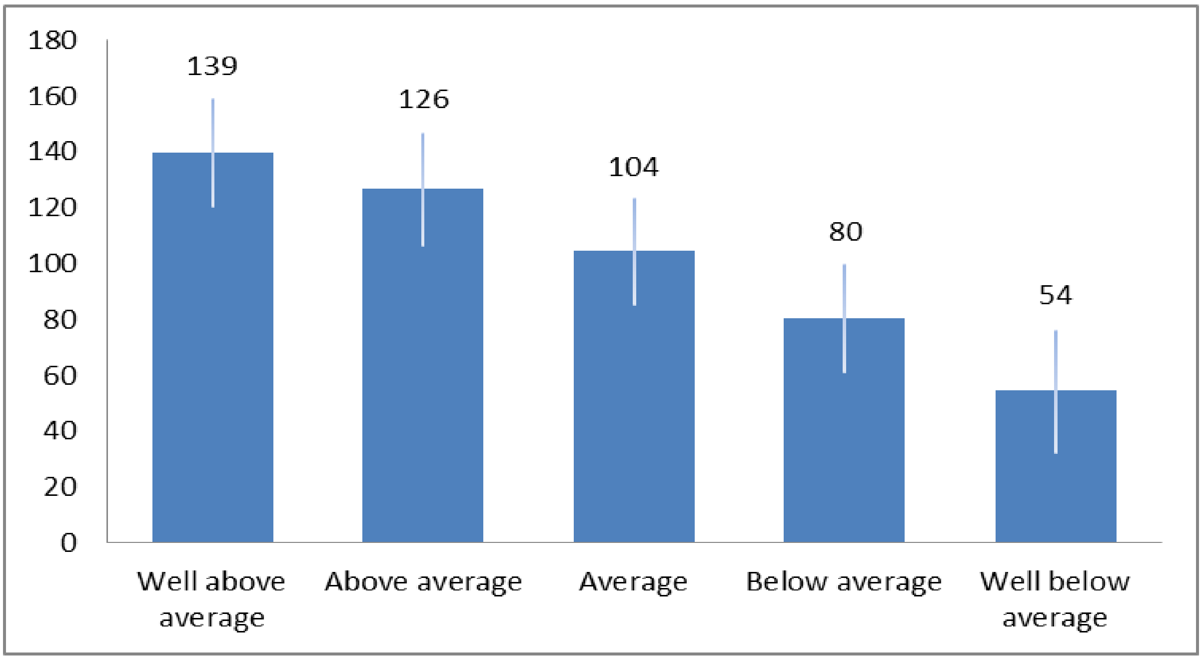
*Individual N in brackets after each characteristic

Modelling

Apparent incongruities between the distribution of pupils with varying characteristics according to Word Reading Score and to teacher judgements of reading 'ability and attainment' begin to support the hypothesis that, as suggested by previous research, there may be biases within teacher perceptions of pupils according to pupil characteristics. In order explicitly to investigate this, regression modelling compares teacher judgements of pupils who differ according to a given characteristic but who score at the same level on the relevant cognitive test.

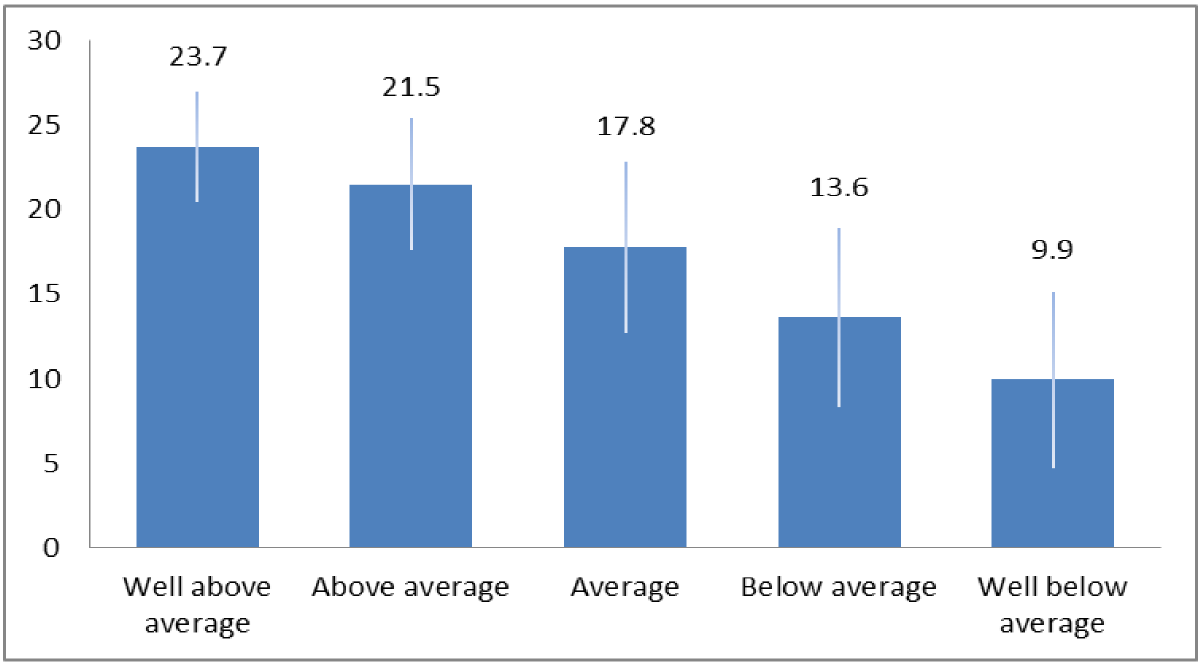
A necessity of the methodology here is the existence, across the sample, of an overall general relationship between performance on each cognitive test and teacher assessment of pupil 'ability and attainment' in the relevant domain. This relationship is strongly apparent. For example, using the whole sample, a naïve regression of BAS Word Reading test score on whether a pupil's teacher perceives their reading as 'above average' indicates that each additional point scored on the Word Reading test (range 10-214) is related to a likelihood of being judged 'above average' increased by 1 percentage point ($p <. 001$). For teacher judgements of reading 'below average', the relationship is inverted and there is a decrease of -1 percentage point ($p < .001$). The relationship between point increase in Progress in Maths score (range 0 – 28) and judgement of 'above average' in maths is 4 percentage points ($p < .001$). For judgements below average it is -3 percentage points ($p < .001$). Figure 6 presents the average BAS Word Reading test score of pupils judged to be at each level of reading 'ability and attainment' by their teacher, and Figure 7 presents the equivalent information for maths scores and judgements. Vertical lines represent one standard deviation on either side of each mean.

Figure 6: Mean BAS Word Reading score of pupils with each teacher judgement of reading 'ability and attainment'



(N = 5098, vertical lines = one standard deviation on either side of each mean)

Figure 7: Mean Progress in Maths score of pupils with each teacher judgement of maths 'ability and attainment'



(N = 5087, vertical lines = one standard deviation on either side of each mean)

Figures 6 and 7 illustrate, across all sample pupils, overall consistent and linear associations between test score and teacher judgement. On average, pupils with a higher cognitive test score are judged to have a higher level of 'ability and attainment' by their teacher.

If there are no biases in teacher judgements according to the pupil characteristics of interest, these associations should not vary by income-level, gender, SEN status, ethnicity, or language. Girls and boys, for example, who score at the same level on the Word Reading test, should have equal probabilities of being judged 'above average' at reading by their teacher.[14]

A linear probability model is used to test whether this is the case.[15] The outcome (for example) is whether a child is judged 'above average' at reading, and the predictors: pupil gender, and ability score on the BAS Word Reading test. The likelihood of boys being judged 'above average' at reading by their teacher is thereby compared to the likelihood of girls who score at the same level on the test. Analysis takes the following form:

$$\begin{aligned}&Probability\ of\ being\ judged\ \text{`}above\ average\text{'}\ at\ reading\ by\ teacher_{0-1}\\&\quad = Constant + \beta Boy_{0/1} + \beta Performance\ on\ BAS\ word\ reading + error\end{aligned}$$

The coefficient for boys represents the percentage point difference in likelihood, compared to girls who score equivalently on the Word Reading test, of being judged 'above average' at reading by their teacher. A coefficient of *0* would indicate that there is no bias according to gender in teacher assessments of reading ability.

Modelling is repeated separately for each pupil characteristic and outcome, resulting in the following basic analyses (Table 5).

[14] The cognitive test scores themselves are of course subject to error, and it is possible that biases on the part of MCS administrators might influence children's performance (for example, if interviewers encourage some children more than others according to perceptions of different levels of ability). Therefore, an additional assumption of the methodology here is that stereotypes held by interviewers would be in the same direction as those held by teachers – therefore leading to under- rather than over-estimates in findings.

[15] Whether results are affected by choice of model type is tested and reported in Annex C; a binary logistic model produces equivalent findings.

Table 5: Variables used in and structure of basic models

Model	Outcome	Predictors (pupil-level)	
1	Teacher judgement of reading *above* average / not	BAS Word Reading test ability score	+ SEN / not
2			+ above / below 60% income
3			+ English only / additional languages
4			+ gender
5			+ ethnicity
6	Teacher judgement of reading *below* average / not	BAS Word Reading test ability score	+ SEN / not
7			+ above / below 60% income
8			+ English only / additional languages
9			+ gender
10			+ ethnicity
11	Teacher judgement of maths *above* average / not	Progress in Maths score	+ SEN / not
12			+ above / below 60% income
13			+ English only / additional languages
14			+ gender
15			+ ethnicity
16	Teacher judgement of maths *below* average / not	Progress in Maths score	+ SEN / not
17			+ above / below 60% income
18			+ English only / additional languages
19			+ gender
20			+ ethnicity

Results: Biases in teacher perceptions

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at reading

Figure 8 indicates the average extent to which MCS pupils who differ according to each characteristic (SEN / not, income-level, language(s) spoken, gender, ethnicity) are under-assessed at reading, in relation to comparators who score equivalently on the Word Reading test. Findings from five different models are presented together in this chart. As described in the methods section, separate models were estimated for each characteristic.[16] Model coefficients and p-values can be found in Annex D.

Boys, children who speak other languages in addition to English, children from less wealthy families, and children with any recognised diagnosis of SEN appear less likely to be judged 'above average' at reading by their teacher – despite scoring equivalently to their comparison counterparts in the Word Reading test. All these differences are significant at the 5% level.

MCS pupils of all non-White ethnicities also appear less likely to be judged 'above average' at reading (compared to White pupils), and the differences from the White reference group are again significant at the 5% level. The biases portrayed in Figure 8 are therefore largely congruent with patterns of under-attainment reported in DfE statistics (though findings for the Indian group are anomalous).

Returning to the descriptive statistics outlined in Table 3 and Figures 2 and 3, it seems that there may feasibly be a number of potential mechanisms at work here. For example, sample low-income pupils tend generally to be more likely to score at lower levels on the BAS Word Reading test. But Figure 8 indicates that *all* low-income pupils are, on average, less likely than their higher-income counterparts who attain equivalent test scores to be judged highly by their teacher. In line with the theory that teachers are stereotyping pupils according to preconceived evaluative templates which have some kind of empirical grounding, this provides a tentative suggestion that teachers may be generalising from average performance at the group level to a biased perception of pupil performance at the individual level (as proposed by e.g. Burgess and Greaves, 2009) – an inaccurate process, which appears to downgrade able pupils from less wealthy families. [17]

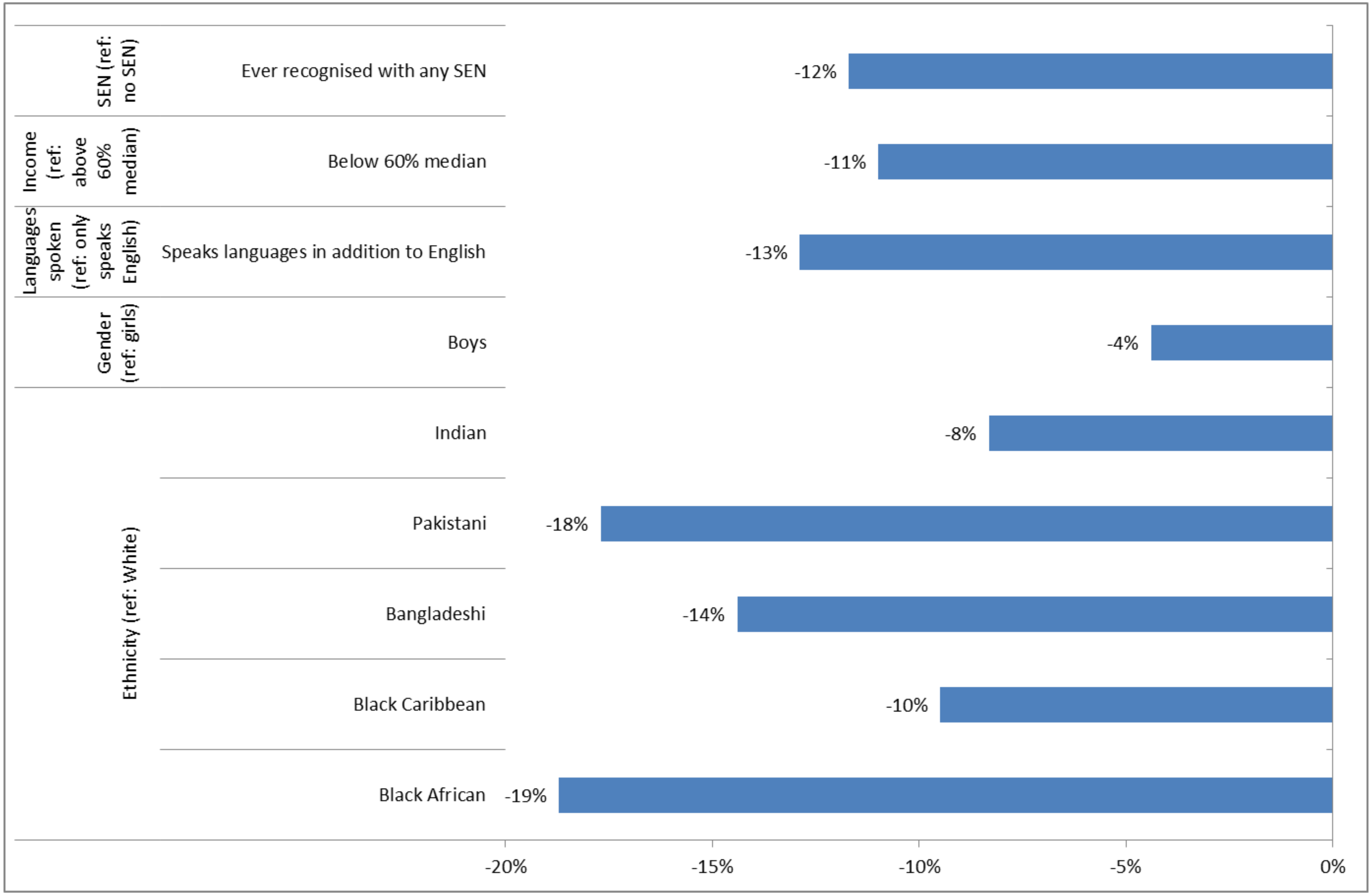
However, in contrast, and directly challenging this explanation, Figure 3 also showed that sample Bangladeshi, Black African, Indian and EAL pupils tended to score highly on the Word Reading test – while Figure 8 indicates that these groups tend to be under-assessed. If the Word Reading test is a reasonable indicator of manifest pupil performance, this raises the question: what exactly is driving this apparent downwards bias in judgement, this

[16] In all models, a one point increase in Word Reading Score is associated with a one percentage point increase in likelihood of being judged 'above average' ($p < .001$).

[17] However, it is important to remember that this is a sample of pupils and teachers which cannot be assumed to exactly replicate the population; therefore performance of pupils within the sample may not exactly replicate teachers' experiences with pupils of different groups (potentially undermining the suggestion that there may be a probabilistic empirical foundation for the stereotyping indicted here).

possible unfavourable stereotyping of these pupils? Might the conclusions of previous research, that there may be, at the institutional level, an unfounded ‘structural inequality’ (Strand et al, 2010) or endemic ‘negative… thinking’ (Maylor et al, 2009) about certain pupil groups go some way towards explaining the biases found here?

Figure 8: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'above average' at reading by their teacher, compared to pupils with the reference characteristic



Ns = (from top to bottom) 5046; 5094; 5098; 5098; 5052
All results significant at the 5% level

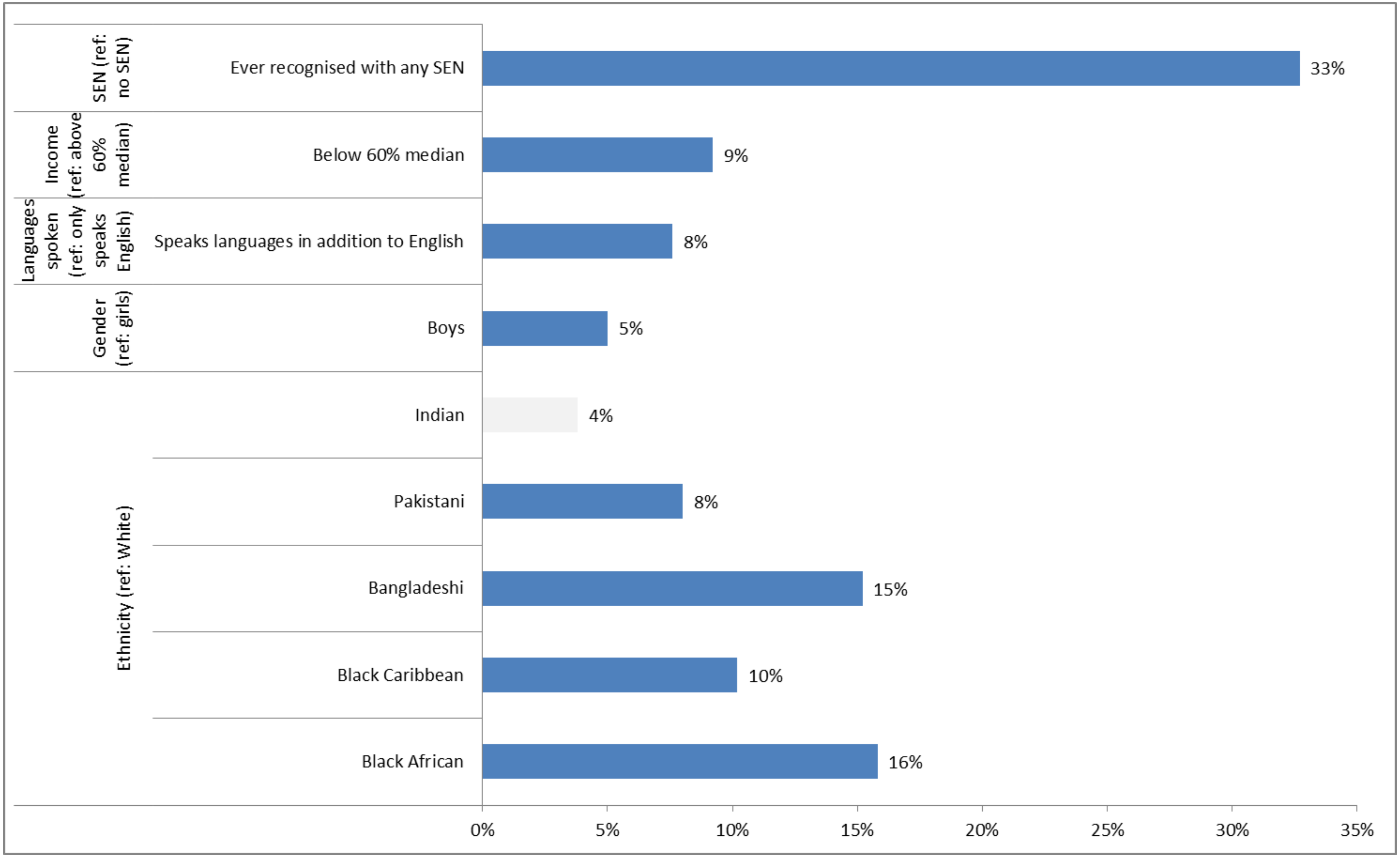
Biases in teacher judgements of whether pupils are of ‘below average’ ability and attainment at reading

Figure 9 continues to indicate the extent to which MCS pupils are under-assessed at reading, compared to peers who score equivalently on the Word Reading ability test, but who differ according to each characteristic.[18]

All differences by pupil characteristic are significant at the 5% level, except the difference between Indian and White pupils. Each difference is congruent with the results above, inverting the findings for teacher judgements of pupils as ‘above average.’ For example, as well as being *less* likely to be evaluated as ‘above average’ than their non-SEN-diagnosed peers scoring equivalently on the Word Reading test, pupils with any diagnosis of SEN are *more* likely to be judged as being of ‘below average’ ability and attainment in reading.

[18] In all models, a higher Word Reading score is associated with a lower predicted probability of being judged ‘below average’ ($p < .001$ in each).

Figure 9: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'below average' at reading by their teacher, compared to pupils with the reference characteristic



Ns = (from top to bottom) 5046; 5094; 5098;
5098; 5052
All results but *Indian* significant at the 5% level

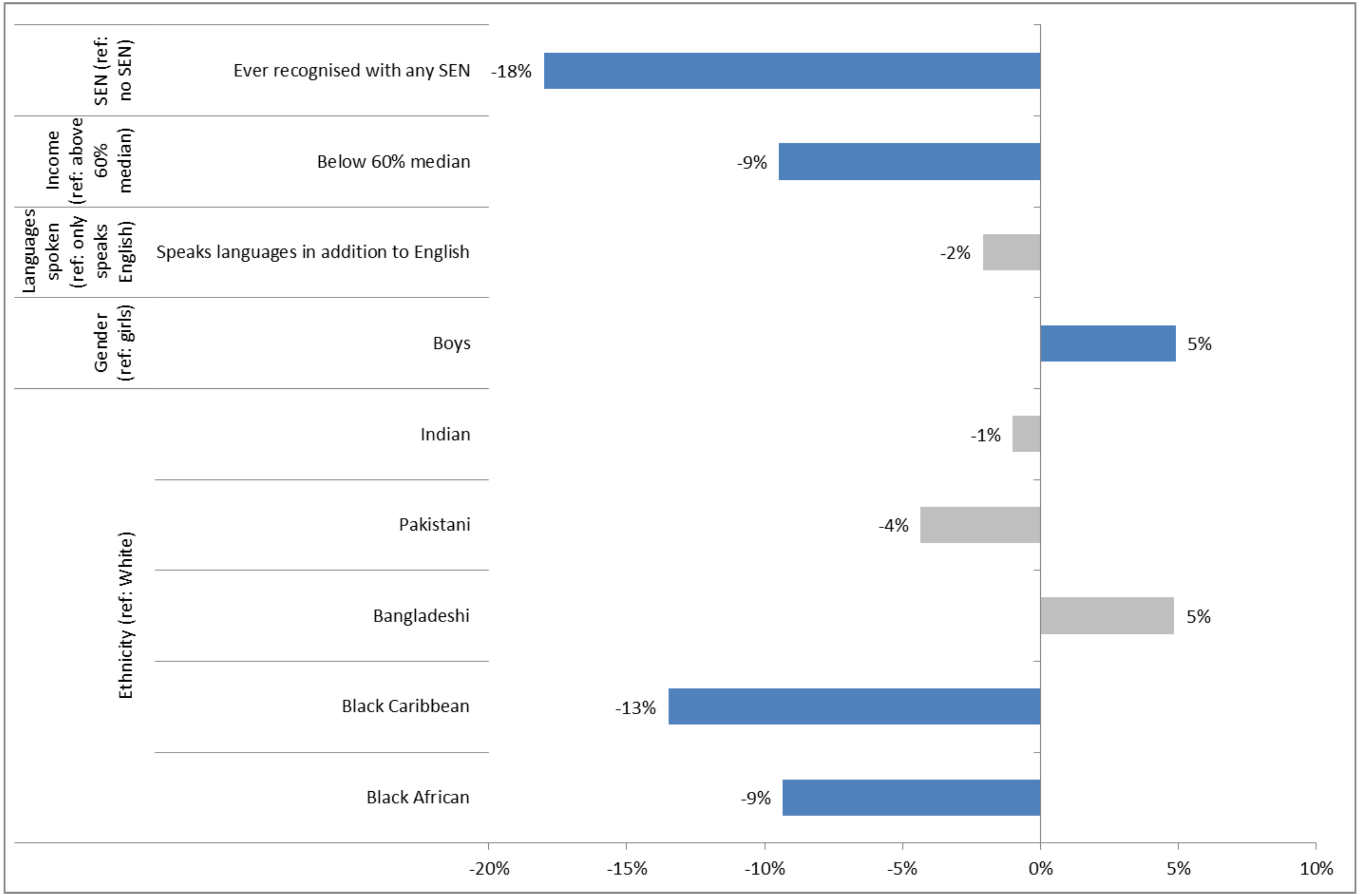
Biases in teacher judgements of whether pupils are of ‘above average’ ability and attainment at maths

In contrast to the apparent biases according to pupil group described for teacher judgements of reading (and in line with a lesser apparent incongruity according to the descriptive statistics), fewer disparities emerge for maths (Figure 10). No significant difference in teacher perceptions is found between MCS pupils speaking only English / speaking an additional language, and pupils of most ethnicities are as likely as White pupils scoring at the same level on the Progress in Maths test to be evaluated as ‘above average.’[19]

Inverting the relationship indicated for judgements of reading ‘above average,’ boys are *more* likely than girls to be judged ‘above average’ at maths, while sample Black African and Black Caribbean pupils are notably less likely than their equivalently performing White counterparts to be judged ‘above average’ – along with children from less wealthy families, and those with any recognised SEN.

[19] As expected, a higher Progress in Maths score is significantly related to an increased predicted probability of being judged ‘above average in all models ($p < .001$ in each model).

Figure 10: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'above average' at maths by their teacher, compared to pupils with the reference characteristic



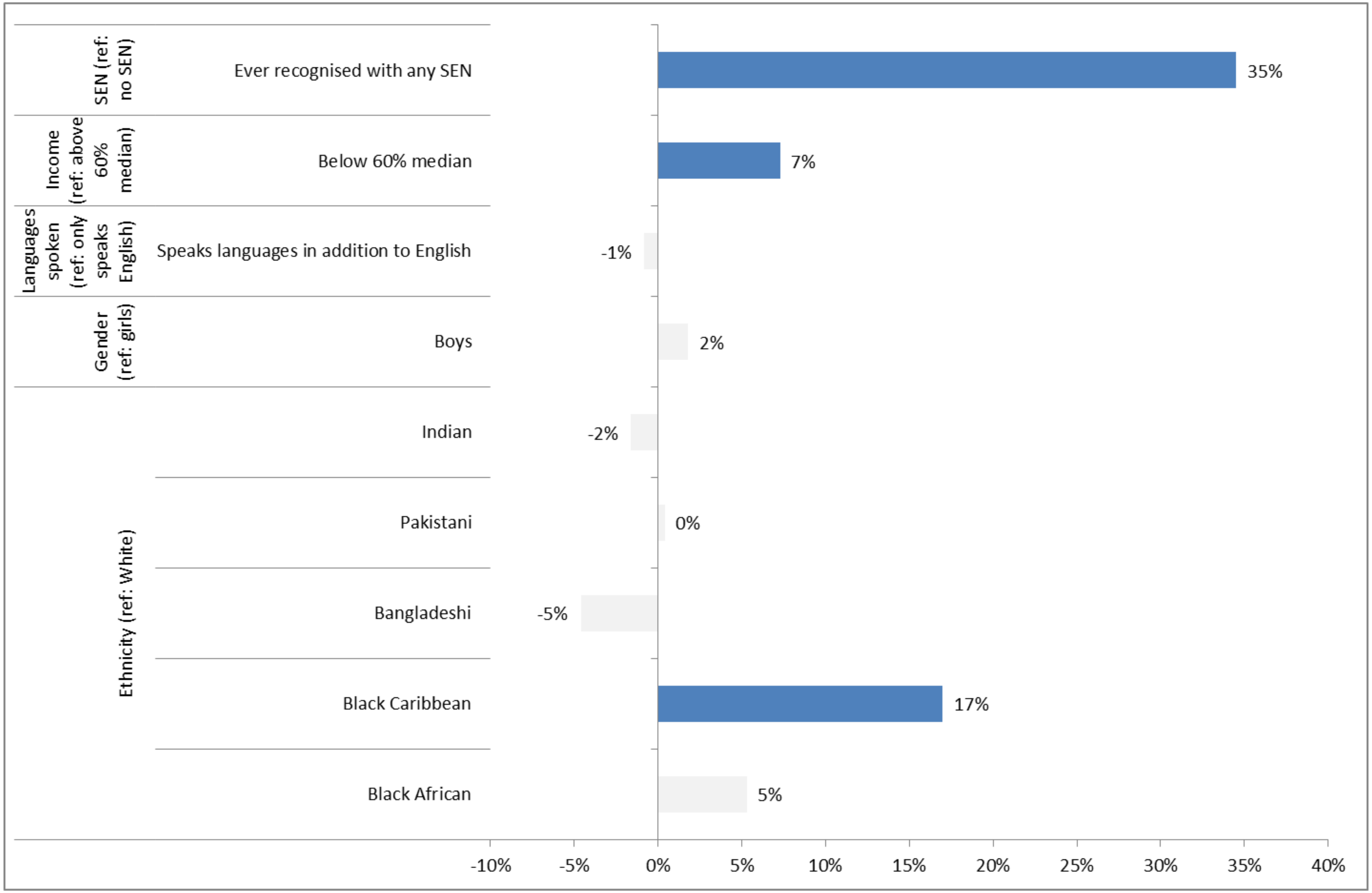
Ns = (from top to bottom) 5034; 5082; 5086;
5086; 5040
Blue bars significant at the 5% level

Biases in teacher judgements of whether pupils are of ‘below average’ ability and attainment at maths

Again, fewer significant biases are found in tendencies of teachers to rate MCS pupils ‘below average’ at maths – but those that emerge are stark. In line with the trend across reading and maths, pupils from less wealthy families are more likely to be under-assessed compared to equivalently scoring, more wealthy counterparts. Pupils who have any diagnosis of SEN are 35 percentage points more likely to be judged ‘below average’ at maths by their teacher, and Black Caribbean pupils 17 percentage points more likely (Figure 11) than equally performing sample peers.[20]

[20] A higher Progress in Maths Score is related to a lower probability of being judged ‘below average’ at maths in all models (all $p < .001$).

Figure 11: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'below average' at maths by their teacher, compared to pupils with the reference characteristic



Ns = (from top to bottom) 5034; 5082; 5086;
5086; 5040
Blue bars significant at the 5% level

Differences in bias according to cognitive test score level

In order to investigate whether the biases demonstrated so far exist for pupils scoring across the whole range on the cognitive tests, or whether they are driven by pupils scoring at certain levels, analysis now interacts test score level with pupil characteristic. To avoid assuming that the relationship between test score and extent of bias must be linear, equal quintiles are created, and denote whether a pupil scores in the bottom, second lowest, middle, second highest, or highest group on each test, respectively.

These quintiles are formed of all pupils who have a teacher judgement in the corresponding domain – so, for BAS Word Reading scores, the quintiles are created for 5098 pupils, and have cut-off points of 86, 103, 117, and 134 (scores range from 10 to 214). Progress in Maths score levels are allocated according to cut-off points of 12.8, 18, 21, and 24 (scores range from 0 to 28), for 5086 pupils.

Teacher judgement is now predicted by respective pupil characteristic and the categorical test score level variable, and the two predictors are interacted. Predicted probabilities are produced for pupils with the given characteristic who score at the level of interest – for example, for boys scoring in the bottom quintile and for girls scoring in the bottom quintile. The magnitude and the significance of the mean difference between boys and girls who score at this level is estimated. As in the previous section, analyses are performed separately for gender, for special educational needs status, for ethnicity, for income-level, and for language(s) spoken.

Tables 6 to 9 present the differences in means for pupils who score within the same quintile on the relevant cognitive test, but who differ according to each characteristic. Asterisks denote differences significant at $p < .05$.

Table 6: Teacher judgements of reading 'above average:' Difference in mean predicted probabilities for pupils with each characteristic, within each Word Reading test score quintile.

	Bottom	Second lowest	Middle	Second highest	Top
SEN (difference from no SEN)	-3	-17*	-31*	-26*	-8
n = 5046					
Below 60% median income (difference from above)	-2	-7*	-.21*	-.20*	-13*
n = 5094					
Additional languages (difference from English only)	0	-11*	-20*	-14*	-12*
n = 5098					
Boys (difference from girls)	-1	-6*	-5*	-8*	-2
n = 5098					
Indian	4	-8	-17*	-12	1
Pakistani	-1	-17*	-23*	-27*	-12*
Bangladeshi	-1	-4	-30*	6	-21*
Black Caribbean	-1	-11	-22*	-14	-6
Black African	-1	-7	-41*	-16*	-18*
(all: difference from White); n = 5052					

* = p < .05

Table 6 indicates that the bias in teacher judgements of SEN pupils is driven primarily by evaluations of pupils scoring around the mean. Pupils with SEN who score in the middle quintile on the Word Reading test are 31 percentage points less likely to be judged 'above average' by their teacher than pupils scoring similarly in the middle quintile with no SEN diagnosis. In contrast, there is a difference of only 3 percentage points for pupils with and without SEN scoring in the bottom quintile. (There is of course less scope for possible variation among pupils in the bottom quintile, given the generally low predicted probabilities of all these pupils and the lower-bound cut-off point of 0).

Biases according to gender appear to be stronger for those pupils who score around the average, and are not significant for pupils in the top and bottom quintiles. In contrast, biases in judgements of low-income pupils are apparent at all test score levels except the lowest: they are large both at the mean and for higher-scoring, low-income children, who have a lower probability than their equivalently high-scoring but wealthier peers of being judged 'above average.' Similarly, pupils speaking languages in addition to English appear penalised at every test score level except the bottom – with a peak for those pupils scoring in the middle quintile, who are 20 percentage points less likely than their peers who speak only English to be judged 'above average.'

Among Indian and Black Caribbean pupils, Table 6 indicates that it is only middle scorers who are under-assessed at reading by their teacher; the probabilities of higher and lower-scorers are not significantly different from equivalently performing White peers. A similar pattern emerges for Bangladeshi pupils, save for a spike in apparent bias among those in the top quintile, who, despite scoring at the highest level on the reading test, are 21 percentage points less likely to be judged 'above average' at reading than White pupils performing at the same level. In contrast, Pakistani pupils scoring at all levels save the bottom are subject to apparent under-assessment, while it is average and higher-scoring Black African pupils who are much less likely than White children to be judged 'above average.'

Table 7 presents an overall tendency for biases in teacher judgements of pupils as 'below average' at reading to largely be driven by lower scoring pupils, in the bottom two quintiles. Boys who score in the bottom quintile, for example, appear to face a double penalty: as well as attaining a low score on the Word Reading test, they are more likely than equivalently scoring girls to be judged 'below average' at reading by their teacher.

Not all the biases in judgements of pupils as 'below average' at reading are explained by pupils scoring at the lower end, however. Pupils with SEN scoring at any level are more likely to be judged 'below average,' and Bangladeshi and Black African pupils scoring around the average, in the middle quintile, are much more likely than similarly scoring White peers to be downwardly assessed.

Table 7: Teacher judgements of reading 'below average': Difference in mean predicted probabilities for pupils with each characteristic, within each Word Reading test score quintile.

	Bottom	Second lowest	Middle	Second highest	Top
SEN (difference from no SEN)	39*	35*	19*	8*	1*
n = 5046					
Below 60% median income (difference from above)	15*	10*	8*	3	0
n = 5094					
Additional languages (difference from English only)	7*	16*	7*	1	0
n = 5098					
Boys (difference from girls)	11*	5*	0	0	0
n = 5098					
Indian	-8	7	4	2	0
Pakistani	10*	13*	9*	2	0
Bangladeshi	18*	19*	28*	-1	3
Black Caribbean	2	32*	11	-1	0
Black African	20*	32*	18*	2	0
(all: difference from White); n = 5052					

* = $p < .05$

Table 8 indicates that it is particularly higher scoring pupils with SEN and girls who are less likely to be judged ‘above average’ at maths by their teacher. Though they attain high levels on the Progress in Maths cognitive test, these children are less likely than counterparts who score at the same level to be evaluated as of relatively superior ability and attainment. Pakistani and Black Caribbean pupils who score in the middle quintile are less likely to be judged ‘above average’ than similarly scoring White pupils – but it is top scoring Black African pupils for whom bias in teacher judgement is most pronounced. Despite scoring at this top level, Black African children are 39 percentage points less likely than top scoring White pupils to be judged ‘above average.’

Table 8: Teacher judgements of maths ‘above average:’ Difference in mean predicted probabilities for pupils with each characteristic, within each Progress in Maths test score quintile.

	Bottom	Second lowest	Middle	Second highest	Top
SEN (difference from no SEN)	-10*	-16*	-27*	-32*	-28*
n = 5034					
Below 60% median income (difference from above)	-3	-5	-19*	-12*	-18*
n = 5082					
Additional languages (difference from English only)	1	0	-6	-5	-8
n = 5086					
Boys (difference from girls)	-2	7*	5	4	10*
n = 5086					
Indian	0	4	2	-2	-3
Pakistani	-1	-4	-19*	-9	2
Bangladeshi	4	8	-6	5	1
Black Caribbean	-5	-21	-32*	-4	-12
Black African	-2	2	-17	-2	-39*
(all: difference from White); n = 5040					

* = p < .05

Table 9 suggests once more that pupils with a SEN diagnosis who score at any level on the Progress in Maths test are subject to apparent bias in increased tendency of their teacher to judge them ‘below average’ at maths– though it is among lower-scoring children that differences are most evident. Similarly, lower-scoring boys, pupils speaking languages in addition to English, and Black Caribbean pupils appear subject to the potentially compounding double penalty of relatively low test performance and bias in teacher assessments. Low-income pupils scoring at all levels except the top appear to be relatively more likely than higher income pupils to be judged ‘below average.’

Table 9: Teacher judgements of maths ‘below average:’ Difference in mean predicted probabilities for pupils with each characteristic, within each Progress in Maths test score quintile.

	Bottom	Second lowest	Middle	Second highest	Top
SEN (difference from no SEN)	38*	42*	27*	28*	23*
n = 5034					
Below 60% median income (difference from above)	6*	9*	12*	6*	4
n = 5082					
Additional languages (difference from English only)	-7*	0	3	0	1
n = 5086					
Boys (difference from girls)	7*	0	0	0	0
n = 5086					
Indian	-5	-7	-4	-6	0
Pakistani	-5	4	9	2	-3
Bangladeshi	-7	-3	-1	-6	-3
Black Caribbean	15	36*	20	3	4
Black African	4	2	2	11	5
(all: difference from White); n = 5040					

* = $p < .05$

Area-level controls

As the MCS children were selected at wave one through a stratified, disproportionate cluster design (Plewis, 2007) which oversampled areas with high numbers of minority ethnic families and high levels of deprivation, this creates a situation where the pupils in the wave four teacher sample are unevenly spread across the country. Sample pupils with particular characteristics are concentrated in certain areas. For example, children from some minority ethnic groups, and those who speak languages in addition to English are clustered. Lower-income pupils are also disproportionately represented in some regions.

See Annex E for comparisons by government office region (GOR) at wave four interview between sample pupils and primary school pupils in England, at around the time the survey took place.[21] These comparisons roughly examine the extent to which there is disproportionality by region. Particularly, for example, Yorkshire-born Pakistani pupils and lower-income pupils appear proportionally more prevalent in the sample than in the schools population. The DCSF statistics also show some inequality according to region in SEN diagnosis rates, and these patterns do not map directly on to those found in the sample.

It is possible, therefore, that the biases illustrated in the preceding results may arise from variation across local practices, tendencies and perceptions – rather than from homogeneous stereotypes relating to the given pupil characteristics at the level of the teaching profession, as initially hypothesised. To test whether this is the case, analysis adds controls for GOR – for example:

$$\begin{aligned} &Probability\ of\ being\ judged\ \text{‘}above\ average\text{’}\ at\ reading\ by\ teacher_{0-1} \\ &\qquad = Constant + \beta Boy_{0/1} + \beta Performance\ on\ BAS\ word\ reading + \beta GOR + error \end{aligned}$$

In models investigating bias according to SEN status, income-level, and gender, the GOR control comprises respective indicators for each of nine unique regions (North East, North West, Yorkshire and The Humber, East Midlands, West Midlands, East of England, London, South East, South West). In models for ethnicity, regions are collapsed into four super-GORs to ensure representation of children of each ethnic group in every area as categorised, and in models looking at language, seven areas are used (the North East and North West are combined, as well as the South East and South West). See Annex A for further detail on area categorisations.

Because children of non-White ethnicities and children speaking languages in addition to English are not represented / severely underrepresented in some areas, necessitating this amalgamation of regions, analysis is limited and uncertain in the extent to which it can indicate the national situation according to ethnicity and according to EAL. However, children of both genders, of different income-levels, and with / without a SEN diagnosis are well represented in each region – meaning that modelling according to these characteristics can reasonably investigate the hypothesis that stereotypes are manifest nationally.

[21] Using DCSF statistics: DCSF, 2009a; DCSF, 2009b.

As well as regional drivers, it is possible that the biases found in earlier analysis may largely be attributable to teachers of pupils in relatively homogeneous, more wealthy areas, where there are – for example – fewer minority ethnic pupils. In more diverse areas, where teachers have a wider experience of pupils with a variety of characteristics, biases may not be so apparent.

The initial sampling strategy for wave one of the MCS involved dividing local wards into three classes: ‘advantaged’ wards, ‘disadvantaged’ wards, and ‘ethnic minority’ wards. Ethnic minority wards had at least 30% non-White residents according to the most recent census; ‘disadvantaged’ wards fell into the poorest 25% in the country according to the child poverty index at the time; advantaged wards comprised the remainder - those not categorised as ethnic minority nor disadvantaged (Plewis, 2007). Table 10 describes proportions of pupils in the main sample used in this paper within each sample strata of origin.

Table 10: Percentage of sample pupils born in each MCS initial wave one sample strata

	Advantaged	**Disadvantaged**	**Ethnic minority**
White	94.9	85.6	22.2
Indian	1.0	1.4	13.9
Pakistani	0.6	4.2	25.4
Bangladeshi	0.2	0.5	9.8
Black Caribbean	0.2	1.6	4.5
Black African	0.3	1.9	9.5
Above 60% median	86.0	65.7	43.2
Below 60% median	14.0	34.3	56.8
Other languages	3.3	8.4	61.4
English only	96.7	91.6	38.6
Boys	49.7	50.6	50.8
Girls	50.3	49.4	49.2
SEN	20.1	25.6	23.0
No SEN	79.9	74.4	77.0

Table 10 shows that the wave four teacher sample of children used in the current analysis is distributed in a way largely congruent with the intentions of the initial sampling: for example, there are proportionally few White children in the ethnic minority strata, and many more higher-income children in the advantaged strata than the disadvantaged or ethnic minority strata.

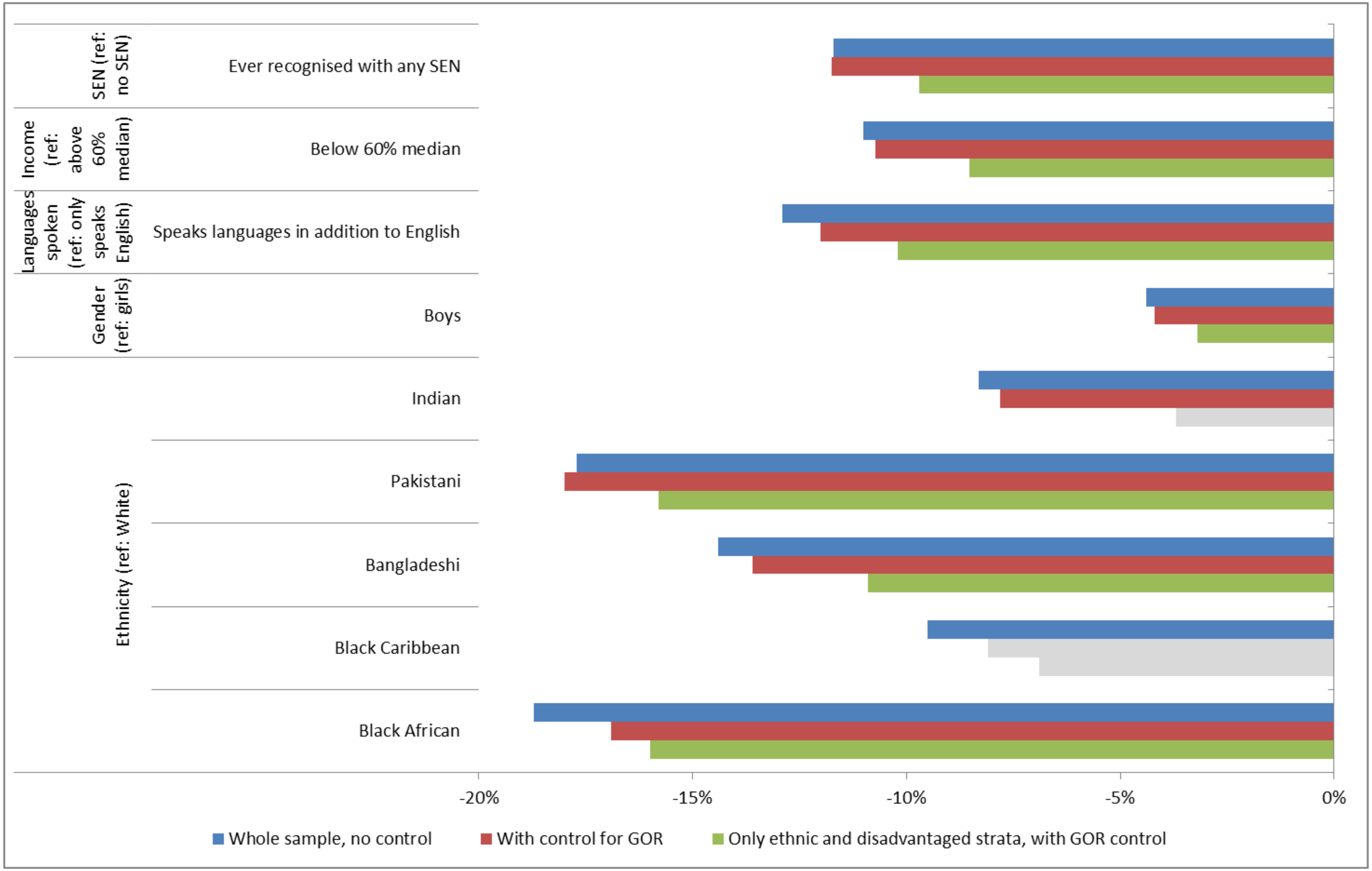
Assuming, therefore, that there is a reasonably strong relationship between the type of area in which a pupil was born and the type of area in which they live at wave four of the MCS, a next stage of analysis goes on to limit the sample to pupils born in disadvantaged and ethnic minority wards only, in order to explore whether the biases indicated thus far are driven by teachers in less diverse areas, or whether they remain when analysis is performed only for pupils in more heterogeneous strata. Controls for GOR continue to be included in these models to account for any other influence of local area.

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at reading, controlled for GOR, then limited to disadvantaged and ethnic minority strata

Most biases in tendencies of teachers to judge pupils as 'above average' at reading continue to be large and significant when analysis controls for GOR and when it is limited to pupils born in more diverse areas (Figure 12). This indicates that the propensities indicated so far are not driven by large-area-specific practices or norms, nor by teachers in more wealthy, homogenous localities, with relatively higher numbers of White pupils. With the exception of the pattern for Indian pupils, tendencies illustrated previously hold among those teachers and pupils in schools with potentially highly varied pupil populations.

However, controlling for GOR renders the bias for Black Caribbean pupils non-significant. Table 13 (Annex E) shows that more Black Caribbean sample pupils live in London than in any other area – and, correspondingly, first stage modelling by ethnicity with the whole, non-restricted sample indicates that pupils living in all other areas are more likely than pupils living in London to be judged 'above average' at reading by their teacher (see Annex D). Apart from this, in all models, most apparent biases according to MCS pupils' characteristics are negligibly lessened by accounting for area or area-type.

Figure 12: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'above average' at reading by their teacher, compared to pupils with the reference characteristic – whole sample, with GOR control, and ethnic and deprived strata pupils only

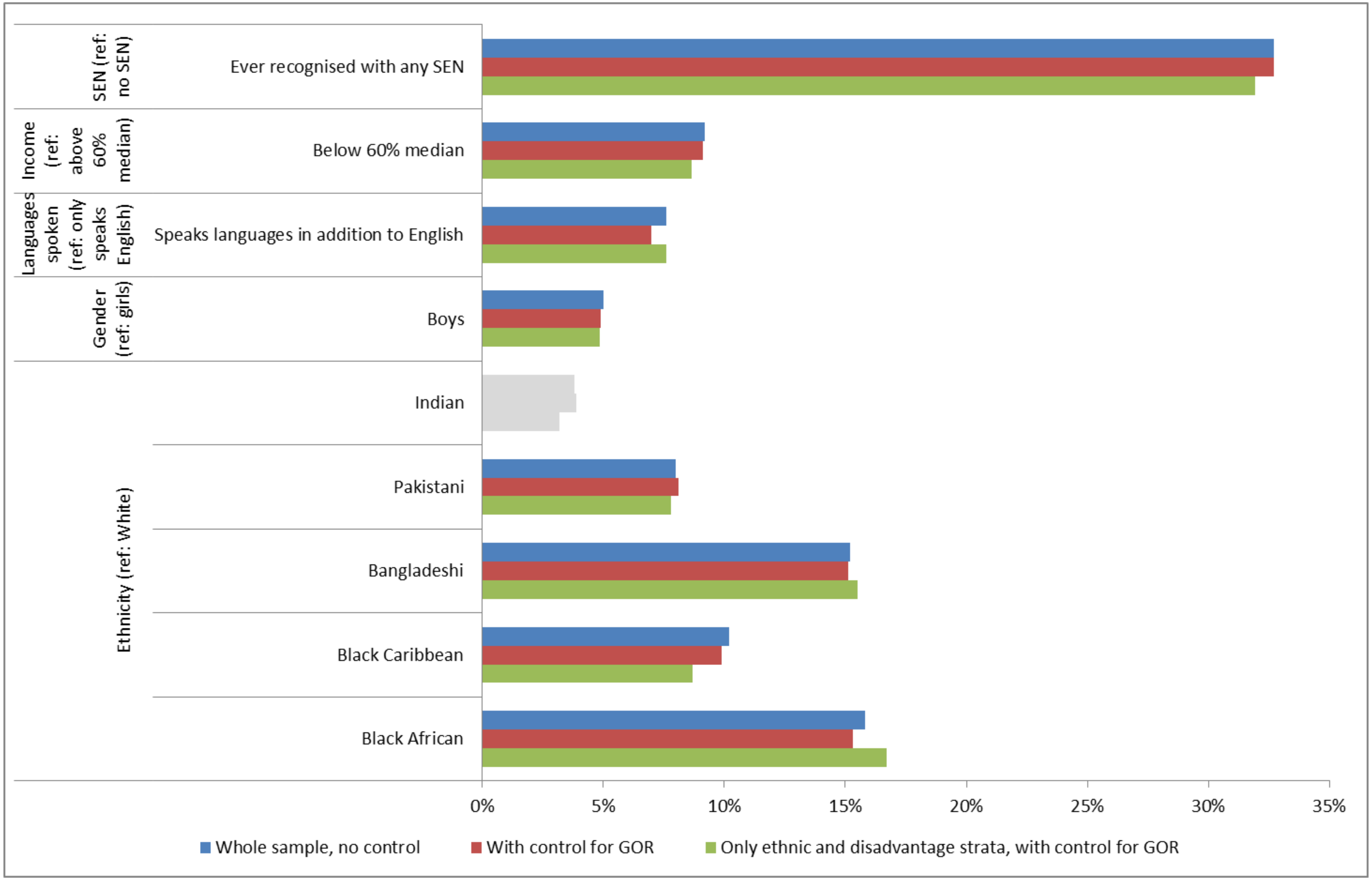


Ns = (from top to bottom) 5046; 5045; 2649; 5094; 5094; 2673; 5098; 5097; 2676; 5098; 5097; 2676; 5052; 5051; 2643). Grey bars NOT significant at the 5% level

Biases in teacher judgements of whether pupils are of ‘below average’ ability and attainment at reading, controlled for GOR, then limited to disadvantaged and ethnic minority strata

The controls here (Figure 13) make less difference still to biases in teacher judgements of pupils as ‘below average’ at reading: all tendencies apparent in the whole sample remain in the analysis controlled for GOR and in the subsample limited to more diverse areas. Coupled with the findings, above, of little mitigation by GOR, or by restricting the sample, of patterns in tendencies of teachers to judge certain pupil groups as ‘above average’ at reading, some support is lent to the hypothesis proposed. In particular, given the reasonable sample distribution according to these characteristics, findings for gender, income-level and SEN status suggest that there may be stereotypes operating nationally, across the level of the teaching profession, which are premised on pupil characteristics.

Figure 13: Difference in percentage point likelihood of pupils with each respective characteristic being judged ‘below average’ at reading by their teacher, compared to pupils with the reference characteristic – whole sample, with GOR control, and ethnic and deprived strata pupils only

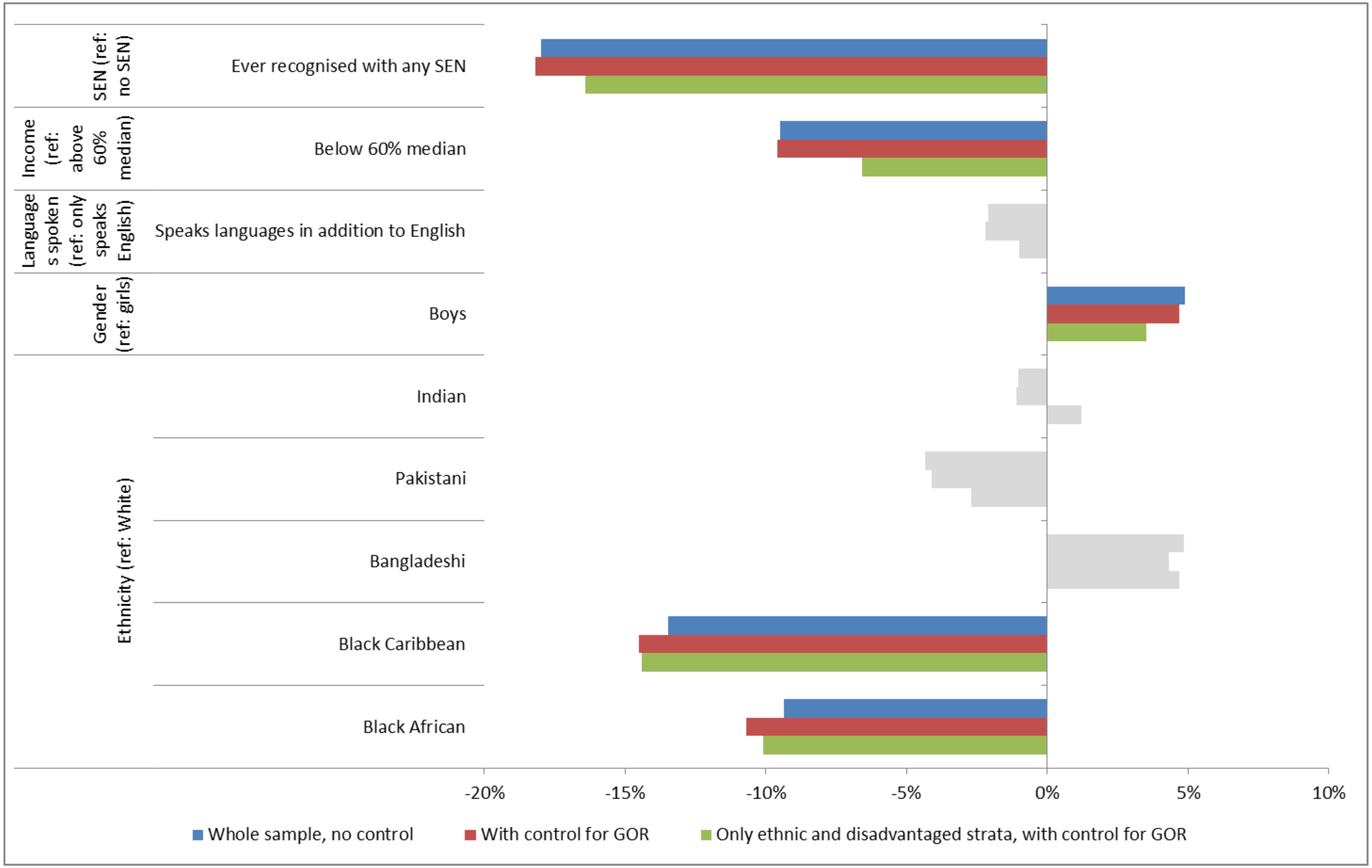


Ns = (from top to bottom) 5046; 5045; 2649; 5094; 5094; 2673; 5098; 5097; 2676; 5098; 5097; 2676; 5052; 5051; 2643). Grey bars NOT significant at the 5% level

Biases in teacher judgements of whether pupils are of ‘above average’ ability and attainment at maths, controlled for GOR, then limited to disadvantaged and ethnic minority strata

As with tendencies to bias in teacher judgements of reading, bias according to pupil characteristic in likelihood of being judged ‘above average’ at maths remain once GOR is controlled for and when the sample is additionally restricted to more diverse areas (Figure 14). A number of GOR areas are significantly related to teacher judgements here at the 5% level in at some models (see Annex D) but, once more, there is support for the possibility that consistent stereotyping of pupils, according particularly to SEN, income-level and gender, is nationally manifest.

Figure 14: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'above average' at maths by their teacher, compared to pupils with the reference characteristic – whole sample, with GOR control, and ethnic and deprived strata pupils only

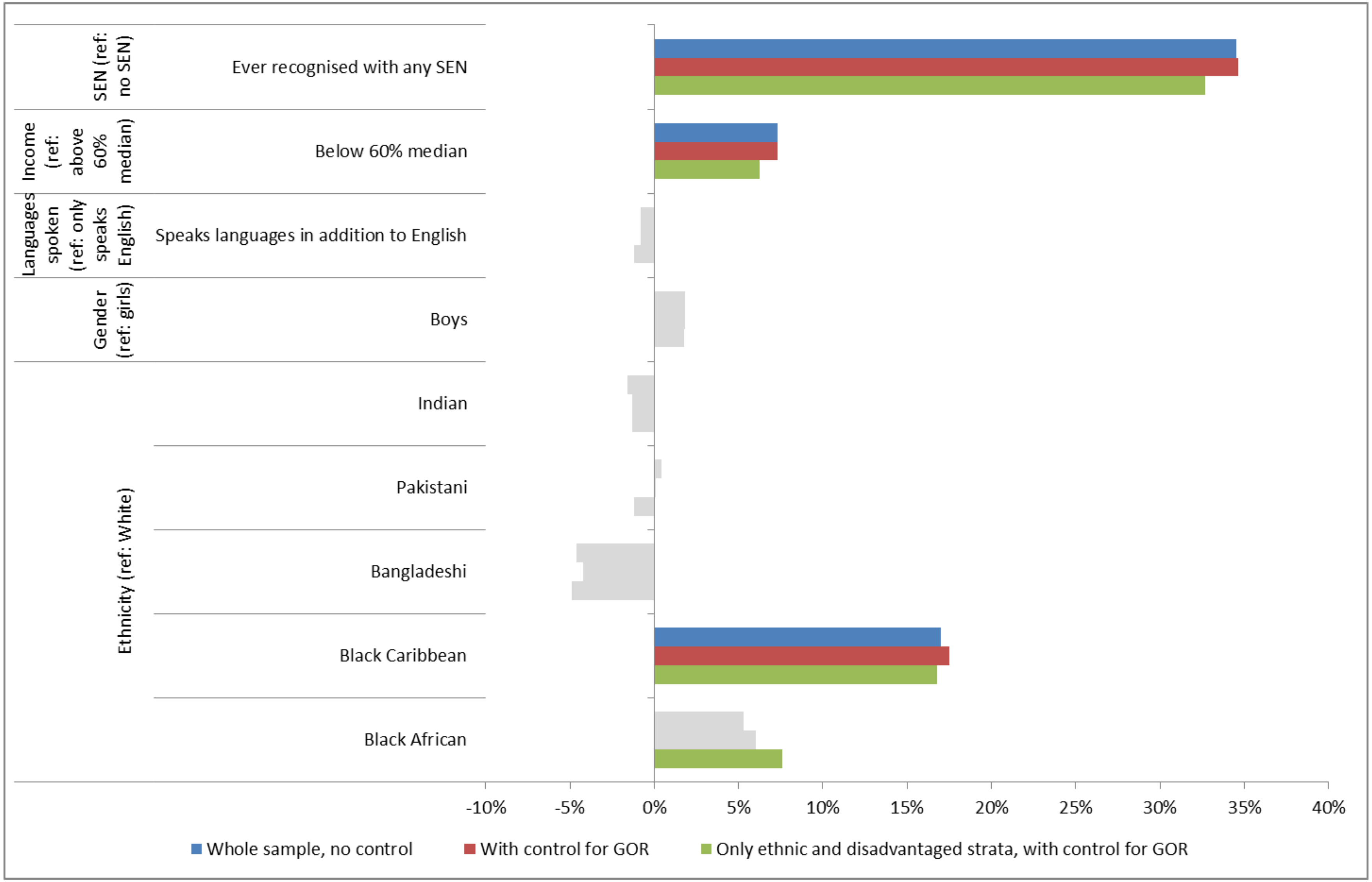


Ns = (from top to bottom) 5034; 5033; 2646; 5082; 5082; 2670; 5086; 5085; 2673; 5086; 5085; 2673; 5040; 5039; 2640. Grey bars NOT significant at the 5% level

Biases in teacher judgements of whether pupils are of ‘below average’ ability and attainment at maths, controlled for GOR, then limited to disadvantaged and ethnic minority strata

All biases in teacher judgements of maths ‘below average’ remain once controls for GOR are added and once analysis is limited to pupils in more diverse areas (Figure 15; see also Annex D) (anomalously, a bias emerges in relation to Black African pupils *only* in the restricted sample).

Figure 15: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'below average' at maths by their teacher, compared to pupils with the reference characteristic – whole sample, with GOR control, and ethnic and deprived strata pupils only



Ns = (from top to bottom) 5034; 5033; 2646; 5082; 5082; 2670; 5086; 5085; 2673; 5086; 5085; 2673; 5040; 5039; 2640. Grey bars NOT significant at the 5% level

Do disparities in income-level explain biases according to gender, SEN, ethnicity or language(s) spoken?

As discussed in the introduction, current government policy emphasises family SES – specifically, income-level – as the key driver of differentiation in pupil attainment, and the crucial characteristic according to which intervention should be targeted. As well as indicating biases according to each other characteristic investigated (gender, SEN, ethnicity and language) analysis so far has, indeed, suggested prevalent, significant biases in teacher judgement according to the income-level of pupils' families.

Firstly, to check and explore whether analysis posited on alternative SES indicators results in a similar, comparable pattern of bias, modelling now substitutes each child's main parent's[22] highest qualification level[23] and each child's main parent's National Statistics Socio-Economic Classification (NS-SEC)[24] (respectively) for income. Mean model-predicted percentage probabilities of attaining each teacher judgement, for pupils whose parent is categorised according to each level, and who score equivalently on the relevant cognitive test, are presented in Tables 11 and 12, below.

Table 11: Mean predicted percentage probability of being judged 'above' / 'below average' for pupils whose main parent is classed as having each level of qualification (controlling for reading / maths test score)

	Probability of being judged 'above average' at reading	Probability of being judged 'below average' at reading	Probability of being judged 'above average' at maths	Probability of being judged 'below average' at maths
No qualifications	37	33	32	31
Overseas qualification only	37	28	36	24
Level one	42	24	36	24
Level two	43	21	37	20
Level three	48	19	41	19
Level four	51	18	45	16
Level five	51	19	47	19

(Ns = reading: 5095; maths: 5083)

In line with findings of bias according to income-level, Table 11 indicates generally that pupils with equal scores on each cognitive test but whose parents have higher qualifications are more likely to be rated highly by their teacher at both reading and maths, and less likely to be rated at a low level.

[22] In most cases, the 'main parent' is the mother.

[23] Wave four variable DMDNVQ00 (this variable is a combination of highest academic and vocational qualifications into overriding categorical levels).

[24] Wave four variable DMD05C00 (this is a 5-category classification of occupation, as per http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html#5)

Similarly, Table 12 suggests that children with parents in managerial and professional occupations are more likely than children with the same test scores who have parents in lower supervisory / technical or routine occupations to be judged 'above average' at reading and maths, and less likely to be judged 'below average.'

Table 12: Mean predicted percentage probability of being judged 'above' / 'below average' for pupils whose main parent is classed as being in each NS-SEC occupational group (controlling for reading / maths test score)

	Probability of being judged 'above average' at reading	Probability of being judged 'below average' at reading	Probability of being judged 'above average' at maths	Probability of being judged 'below average' at maths
Managerial and professional	52	18	47	17
Smaller employer / self-employed	49	20	41	20
Intermediate	47	18	42	17
Lower supervisory / technical	46	18	39	22
Semi-routine / routine	45	22	39	21

(Ns = reading: 4640; maths:)

Given these consistent relationships between family SES and teacher judgements of pupils, it is possible that, as suggested by the 2010 Schools White Paper (see introduction) some of the other biases indicated throughout analysis are driven by differences in SES, as proxied by income – given that, as described in the White Paper, pupils with SEN may disproportionately be lower-income, as well as pupils of some ethnicities [DfE, 2010a]).

In order to test whether level of family wealth appears to dominate and explain bias in teacher perceptions of pupils, modelling now adds income-level (above / below OECD 60% median indicator) as a control to analyses according to each other pupil characteristic. Analysis here continues to control for GOR and to be limited to pupils born in more diverse areas, in order that teachers whose judgements are included will be likely to have experience with pupils of different income-levels and with each of the other characteristics. The (example) model is now, therefore:

$$\begin{aligned} &Probability\ of\ being\ judged\ \text{‘}above\ average\text{’}\ at\ reading\ by\ teacher_{0-1} \\ &\quad = Constant + \beta Boy_{0/1} + \beta low\ income_{0/1} + \beta GOR \\ &\quad + \beta Performance\ on\ BAS\ word\ reading + error \end{aligned}$$

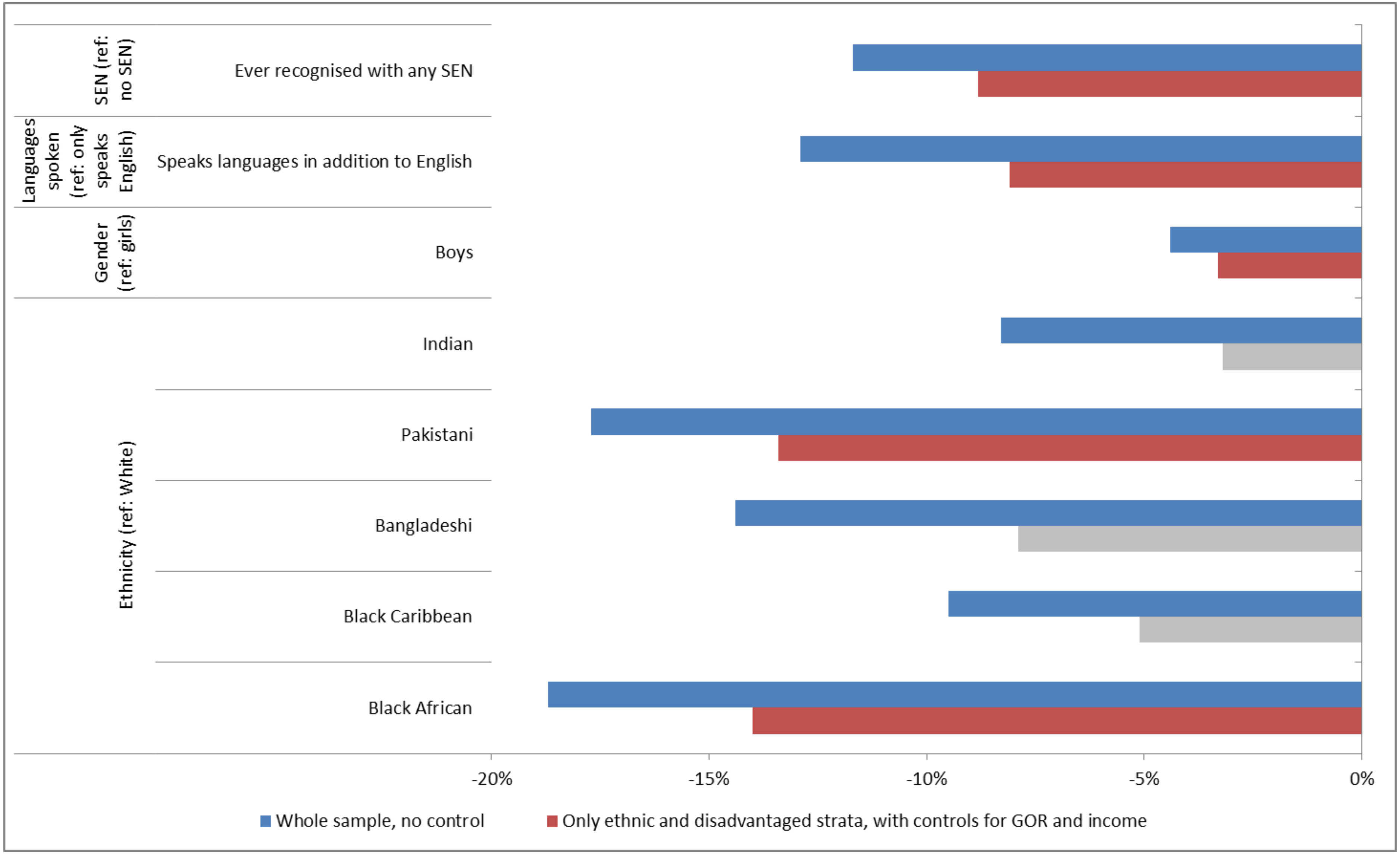
Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at reading, controlled for income and GOR (disadvantaged and ethnic minority strata)

Figure 12 indicated that biases for sample Indian and Black Caribbean pupils are explained through addition of controls for GOR and restriction of analysis to more diverse areas; Figure

16 shows that controlling for income also seems to account for results for Bangladeshi pupils.

However, indications of bias for boys, SEN, EAL, Pakistani, and Black African pupils remain – income does not account for the discrepancies in teacher judgements according to those characteristics. MCS pupils of similar income-levels, who score equivalently on the BAS Word Reading test, but who have a SEN diagnosis, who speak languages in addition to English, who are boys, who are Pakistani, or who are Black African, have a lower likelihood than their counterparts of being judged ‘above average’ at reading by their teacher. (See Annex D for all model coefficients.)

Figure 16: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'above average' at reading by their teacher, compared to pupils with the reference characteristic – whole sample without controls, and ethnic and deprived strata pupils with controls for GOR and income

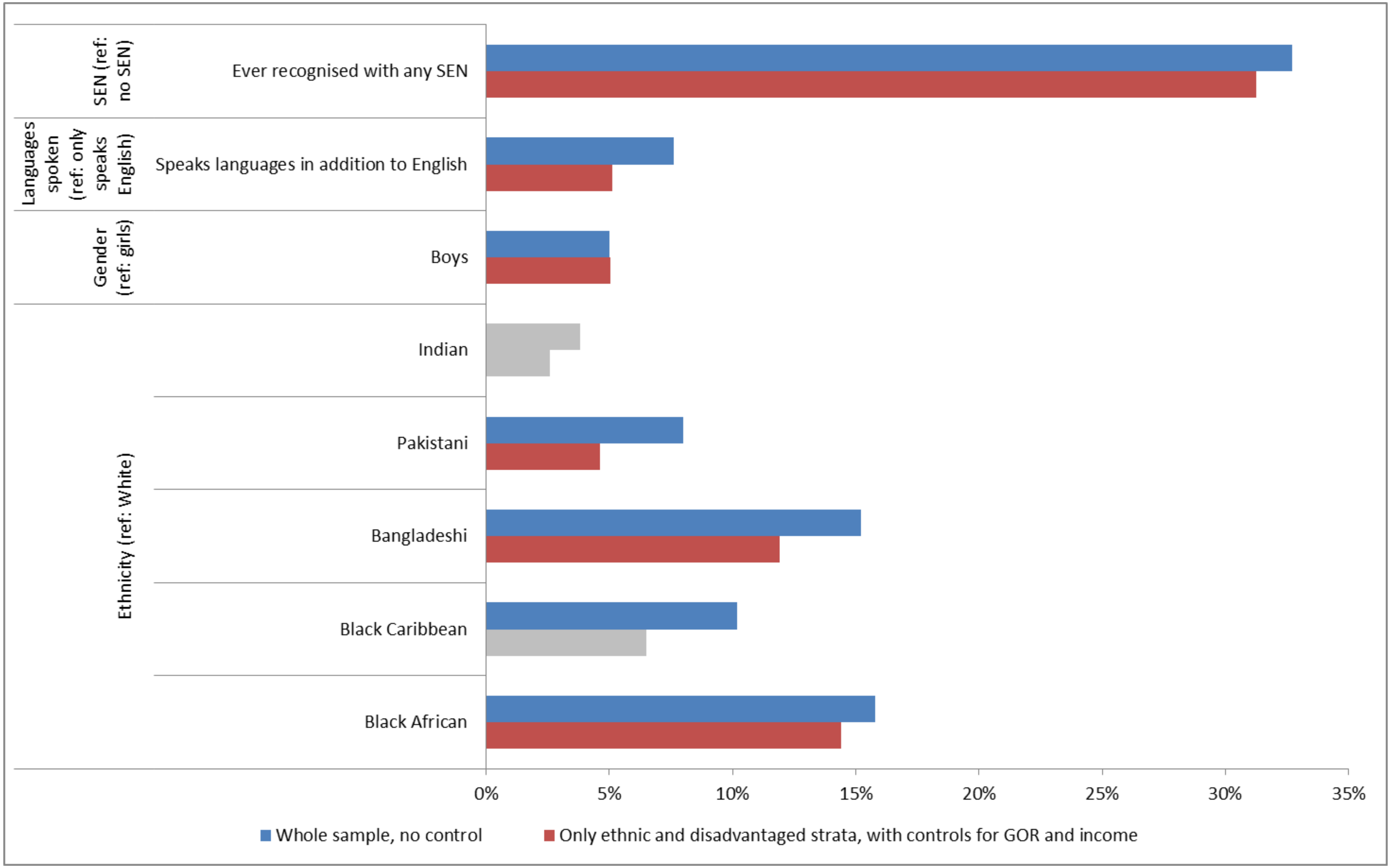


Ns = (from top to bottom) 5034; 2646; 5086; 2673; 5086; 2673; 5040; 2641. Grey bars NOT significant at the 5% level

Biases in teacher judgements of whether pupils are of ‘below average’ ability and attainment at reading, controlled for income and GOR (disadvantaged and ethnic minority strata)

Controlling for income provides no additional explanation for biases in teacher judgements of some MCS pupils as ‘below average’ at reading (Figure 17). Pupils of similar income-levels, who score similarly on the word reading test, but who have a diagnosis of SEN, who speak languages in addition to English, who are boys, or who are Pakistani, Bangladeshi, or Black African, appear more likely to be judged ‘below average’ at reading by their teacher.

Figure 17: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'below average' at reading by their teacher, compared to pupils with the reference characteristic – whole sample without controls, and ethnic and deprived strata pupils with controls for GOR and income

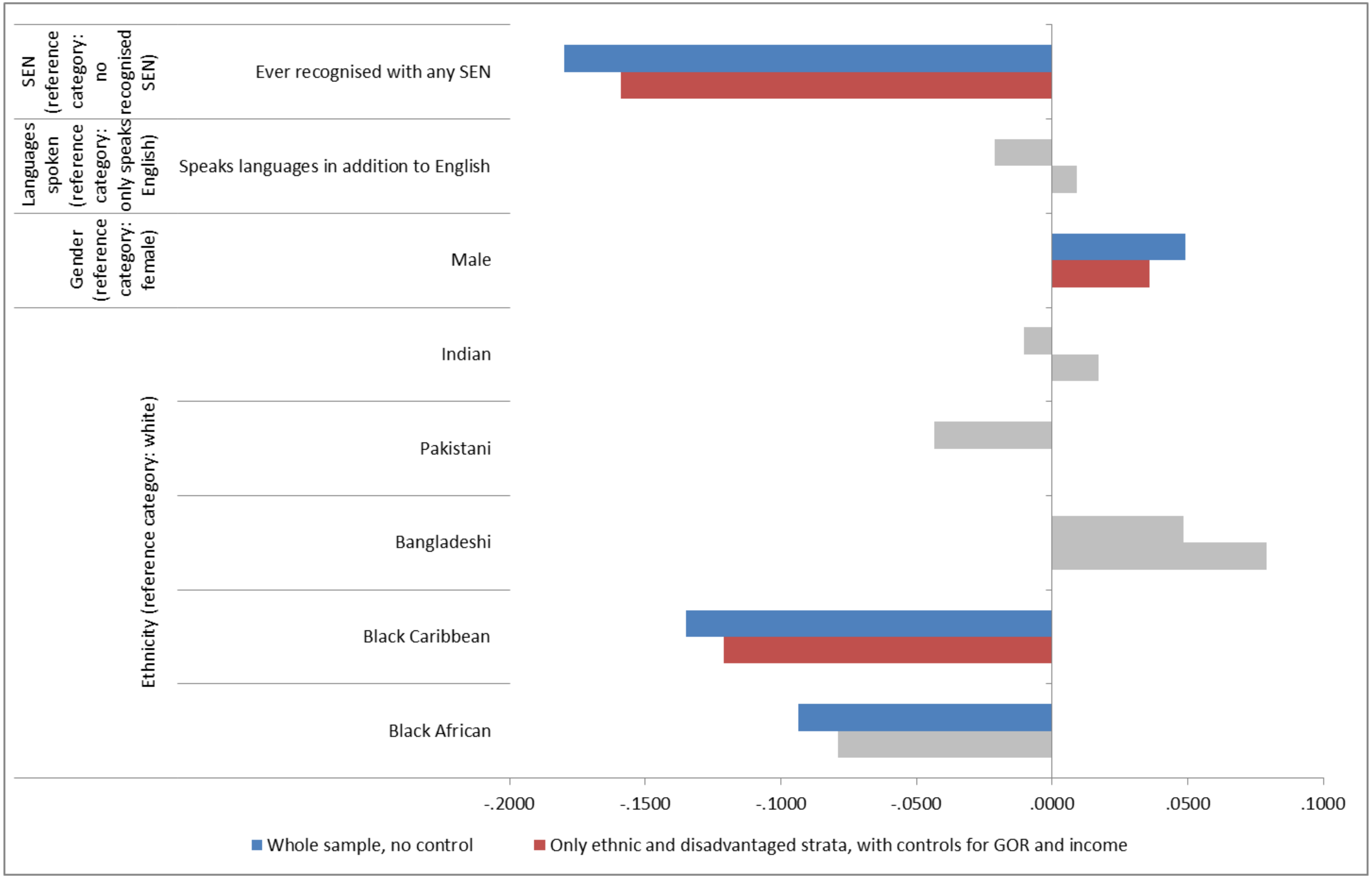


Ns = (from top to bottom) 5034; 2646; 5086; 2673; 5086; 2673; 5040; 2641. Grey bars NOT significant at the 5% level

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at maths, controlled for income and GOR (disadvantaged and ethnic minority strata)

Controlling for family income-level appears to account for some of the relationship for MCS pupils between being of Black African ethnicity and teacher judgement of whether pupils are of 'above average' ability and attainment at maths – but income-level does not explain biases for pupils with SEN, boys, or Black Caribbean pupils (Figure 18). Children with these characteristics who are of similar income-levels and who score equivalently on the Progress in Maths test are less likely to be judged 'above average' at maths by their teacher.

Figure 18: Difference in percentage point likelihood of pupils with each respective characteristic being judged ‘above average’ at maths by their teacher, compared to pupils with the reference characteristic – whole sample without controls, and ethnic and deprived strata pupils with controls for GOR and income

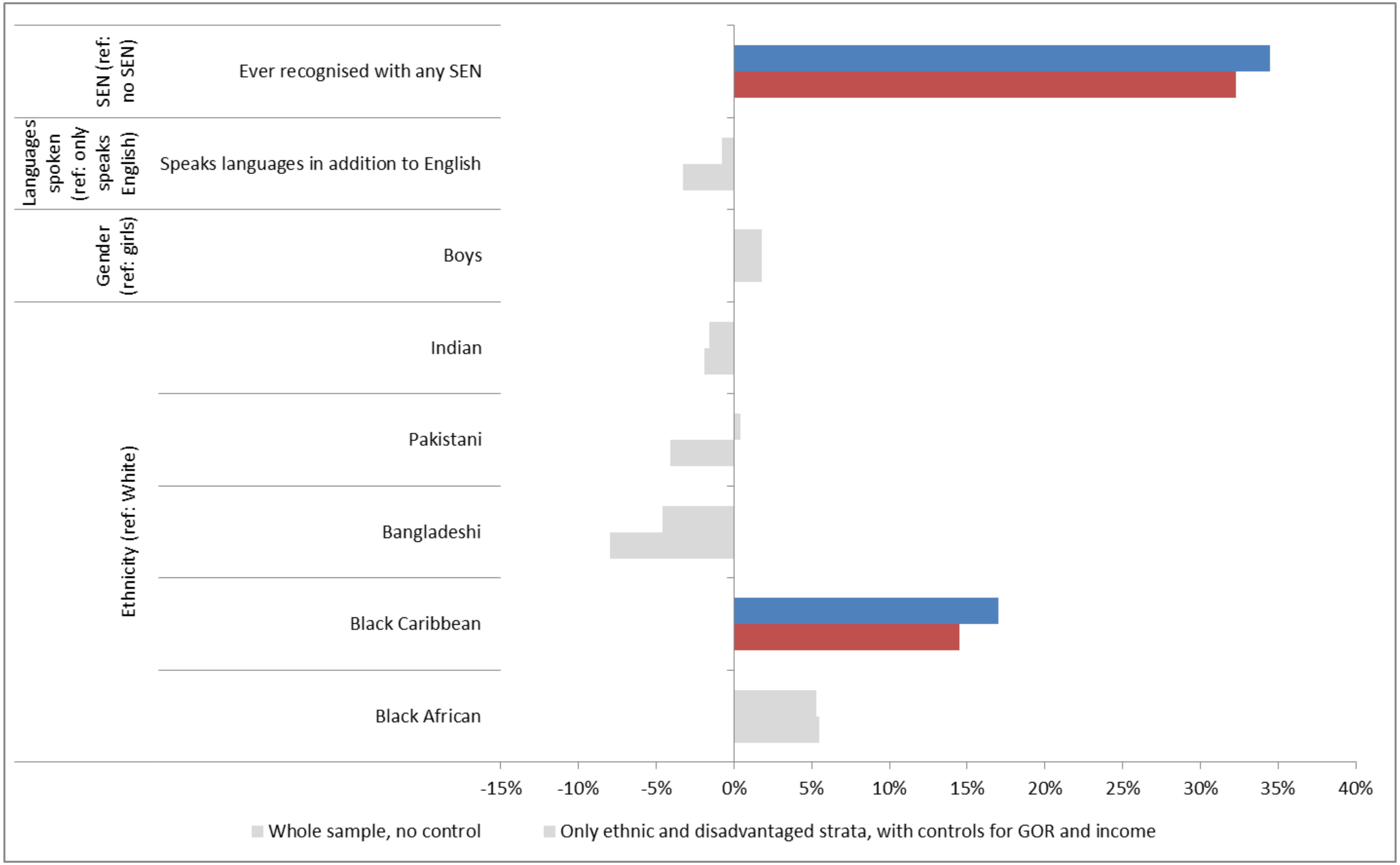


Ns = (from top to bottom) 5034; 2643; 5086; 2670; 5086; 2670; 5040; 2638. Grey bars NOT significant at the 5% level

Biases in teacher judgements of whether pupils are of ‘below average’ ability and attainment at maths, controlled for income and GOR (disadvantaged and ethnic minority strata)

The apparent tendency of teachers disproportionately often to judge MCS pupils who score equivalently on the Progress in Maths test but who have any SEN diagnosis or are of Black Caribbean ethnicity as ‘below average’ at maths remains, even after limiting the sample to more diverse areas and controlling for family income-level (Figure 19).

Figure 19: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'below average' at maths by their teacher, compared to pupils with the reference characteristic – whole sample without controls, and ethnic and deprived strata pupils with controls for GOR and income



Ns = (from top to bottom) , 5034; 2643; 5086; 2670; 5086; 2670; 5040; 2638. Grey bars NOT significant at the 5% level

Summary and discussion

Analysis set out to explore whether there are biases in teacher perceptions of MCS pupils, at age seven, which correspond to the characteristics delineating attainment gaps during primary school (income-level, gender, SEN status, ethnic group, language[s] spoken). By using measures of pupil test performance which are independent of the educational system, and indicators of teacher perceptions of pupils which are not required by or implicit with formal in-school assessments, this paper hoped to investigate the hypothesis that the socio-cognitive process of stereotyping may play a part in creating and sustaining differentiated pupil attainment. Additionally, in line with the current government's focus on poverty as the key driver of educational inequalities, analysis also explored whether family income-level underpins biases according to each of the other pupil characteristics.

Key findings and conclusions

Analysis indicates that:

- At age seven, MCS sample pupils from low-income families are less likely to be judged 'above average' at reading than higher-income peers scoring equivalently on the BAS Word Reading test. Boys, pupils with any SEN diagnosis, and children who speak languages in addition to English are also relatively under-rated by their teachers. In addition, Indian, Pakistani, Bangladeshi, Black Caribbean, and Black African pupils are less likely than their White counterparts to be judged 'above average.' Correspondingly, nearly all of these pupil groups are more likely to be judged 'below average' at reading.

- Low-income MCS pupils, girls, children with any SEN, and pupils of Black African and Black Caribbean ethnicities are also less likely to be judged 'above average' at maths by their teacher. All these children (with the exception of girls and Black African pupils) are, correspondingly, more likely to be judged 'below average' at maths than peers performing at the same level on the Progress in Maths test.

- Neither controlling for local area (GOR) nor limiting the sample to pupils born in areas with higher numbers of lower-income families and more ethnic minority pupils mitigates the vast majority of the biases apparent in teacher judgements. Even in these more heterogeneous localities, bias seems to remain.

- Controlling further for family income-level does not explain the majority of biases according to gender, SEN diagnosis, language, or ethnicity. There appear to be independent biases in teacher perceptions of MCS pupils according to these characteristics, despite pupils scoring equivalently in the relevant cognitive test, and being from families of similar income-levels.

The first overriding research question to be addressed in this paper was:

- At age seven, are there biases in survey-reported teacher judgements of pupils corresponding to each of the key pupil characteristics underpinning recorded primary-age attainment gaps (family income-level, gender, SEN, ethnicity, EAL)?

Analysis has found biases in this sample according to each – and the majority hold both against controls for local area, and when analysis is limited to pupils born in more diverse localities. Indications that bias may occur nationally are strongest in analyses looking at differences according to gender, income-level and SEN status (in MCS sample, pupils of non-White ethnicities and pupils speaking languages in addition to English are not represented / are under-represented in some areas – so the national applicability of results according to ethnicity and EAL cannot be asserted). What is also strongly supported by this paper is the existence of stereotyping according to every characteristic, including ethnicity and language spoken, across both relatively homogeneous areas and much more diverse areas.

Findings using this sample are congruent with previous research indicating over- and under-assessment of pupils according to their characteristics. In addition, because this study uses measures of pupil performance and of teacher perception which are not directly interdependent with or required by the school system, it provides enhanced support for the possibility that teachers, at the national, institutional-level, may stereotype their pupils; that this stereotyping affects the assessment levels awarded to pupils; and that, to some degree, corresponding over and under-assessment may underpin and explain attainment gaps in primary school.

This research also asked:

- Does controlling for income-level explain any biases according to gender, SEN, ethnicity, or EAL?

Most biases in teacher judgements were not accounted for by income-level. Income-level is, in itself, significantly related to biases in teacher judgements of MCS children – but so are gender, SEN status, ethnicity and language spoken. The finding that most other biases remain after income is controlled for implies that an emphasis solely on interventions targeted and distributed according to levels of family wealth is unlikely wholly to eradicate inequalities in primary pupil attainment.

Indeed (and in sympathy with the justification for the recent abolition of CVA scores[25]) findings suggest that explicitly and overtly targeting pupils according to their income-level may go some way towards reinforcing any stereotype that less wealthy pupils should be expected to be less able (though, of course, there are also arguments *for* targeting according to income). Primarily, though, analysis indicates that social and cognitive processes beyond the level of the individual pupil and their family, and outside of the control of the individual pupil or their parents, appear to be at work in shaping pupil trajectories – and that, unless

[25] '…expecting different levels of progress from different groups of pupils on the basis of their…family circumstances…is wrong in principle' (DfE, 2010a).

they are addressed, these processes will continue to play some part in creating and perpetuating inequalities.

Crucially, findings and conclusions in this paper *do not* serve as any condemnation of teachers – as a profession or as individuals – as enacting the process of stereotyping to any unusual (or to any deliberate) degree. As outlined in the introduction, stereotyping is conceived to be a universal, non-conscious, automatic cognitive process which enables speed and efficiency in thought and behaviour. According to theory, all individuals have a propensity to enact the process to some degree: there is no reason that teachers should be exempt, nor unusually prone. Bias in perceptions of pupils is just one manifestation of this human tendency to stereotype. What this paper indicates, simply, is evidence that, at an aggregate level, this human process may be driving some of the inequalities in primary school attainment.

The findings presented here suggest, therefore, that addressing and confronting the process of stereotyping may enable more accurate, more equal, and fairer assessment of pupils.

Tackling stereotyping

It has long been argued that self-awareness and self-reflectiveness are crucial to effective teaching:

> The important thing is for teachers to be explicitly aware of the nature of their responses and how far these are likely to lead to a solution or get in the way of one (Arnold, 1996).
>
> …for teachers to optimise learning they need to have a greater awareness of the complexities of individual differences [and] the important of perceptions and expectations of pupils on learning outcomes…(Hallam and Ireson (1999).

Earp (2010) reviews the cognitive-psychological literature on stereotype activation and consequential behaviours and argues (here, in relation to stereotyping according to ethnicity) once more for mindfulness: 'A teacher who is unaware of the basis for her judgments may conclude that they stem from the realities of her student's performance, rather than (directly or indirectly) from the activation of stereotypes about that student's [ethnic] group.'

Earp goes on to draw from the literature explicit means by which teachers may thwart the automatic cognitive process of stereotyping. Disputing the conclusions reached by researchers such as Bargh (1999),[26] he argues that, although elimination of stereotyping

[26] 'It would be *nice* if stereotypes were found not to be activated automatically. [...] It would be *nice* if, even if automatic activation could not be shown to be prevented, [...] individuals were found to be indeed cognizant of the possibility of being nonconsciously influenced, and when aware of that influence, to have the motivation and the time to effortfully control it. And it would be *nice* if, even if all these propositions failed and stereotypes were shown to be automatically activated and to affect perceptions of and behavior toward a member of a minority group, this influence was still found to be benign. [...] All this would indeed be nice – if it were true. But the relevant research evidence largely contradicts this rosy picture' (Bargh, 1999).

throughout society may well be an infeasible dream, ‘Teachers are just the sort of people who are in a position to automate egalitarian motives’, noting that recent cross-disciplinary studies have indicated that teachers may, with time and effort, ‘train’ and tame the stereotyping mechanism and their own unconscious. This may involve actively learning to draw on alternative stereotypes of pupils, to presume motivation and ability in each student, and consciously to be egalitarian and constructive in feedback to and interactions with pupils. This may ‘reset’ norms and produce new reinforcements, thus breaking down old stereotypes, and eliminating their effect. Earp concludes that, ‘it is essential that schools of education include in their curricula state-of-the-science resources on the unconscious nature of prejudice and the corresponding implications for [the] classroom.’

Initial teacher training content varies according to institution, and continued professional development (CPD) resources are also disparate. Undoubtedly some exist on the importance of teacher perceptions and self-awareness. For example, (2005) CPD materials produced by Oxford Brookes University state that:

> What teachers think influences what teachers do. Teachers’ beliefs and expectations therefore have a significant influence on the quality of their pupils’ learning. Their assumptions, which indeed they may not be conscious they hold, relate not only to how children behave and learn but also to the nature and potential of certain groups of pupils. Teachers’ self-awareness thus becomes a matter of crucial importance in getting the best possible achievement from all pupils…particularly relevant where teachers’ expectations may unwittingly put an artificial ceiling on pupils’ learning or, in some cases, result in their underachieving.

The recommendation of this paper is not, therefore, that the wheel be reinvented. However, findings here suggest that efforts – made through teacher practice, training and professional development, policy-making, and research and evidence-building – to ensure parity, equality and meritocracy in the education system have not yet resulted in a parity of assessment and judgement. Consequentially, this paper recommends that renewed and increased credibility and importance be given to the evidence that biased judgements and stereotyping may be impacting upon and shaping pupil trajectories and attainment, and that resources be directed to enabling and encouraging teachers (and other workers in education) to recognise and challenge the existence and effects of stereotyping – in order that the resulting inequalities may be alleviated.

References

Arnold, R. (1996) *Raising Levels of Achievement in Boys.* Slough: National Foundation for Educational Research.

Bargh (1999). *The cognitive monster: The case against controllability of automatic stereotype effects*. In Chaiken and Trope (Eds.) *Dual process theories in social psychology*. New York: Guilford.

Bradbury (2011a) *Equity, ethnicity and the hidden dangers of 'contextual' measures of school performance.* Race Ethnicity and Education, 14:3, p 277-291

Bradbury (2011b) *Rethinking assessment and inequality: The production of disparities in attainment in early years education.* Journal of Education Policy, 26:5, p 655-676

Burgess and Greaves (2009) *Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities*. University of Bristol: Centre for Market and Public Organisation, Working Paper no. 09/221 http://www.bris.ac.uk/cmpo/publications/papers/2009/wp221.pdf

Chaplin Gray, Gatenby, Simmons, and Huang (2010). *Millennium Cohort Study Sweep Four Technical Report.* http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=844&itemtype=document

DCSF (2007) A Technical Guide to Contextual Value Added 2007 Model http://www.education.gov.uk/performancetables/schools_07/2007GuidetoCVA.pdf

DCSF (2008) 21st Century Schools: A World Class Education for Every Child https://www.education.gov.uk/publications/eOrderingDownload/DCSF-01044-2008.pdf

DCSF (2009a) Schools, Pupils and Their Characteristics: January 2009 http://www.education.gov.uk/rsgateway/DB/SFR/s000843/index.shtml

DCSF (2009b) Special Educational Needs in England: January 2009http://www.education.gov.uk/rsgateway/DB/SFR/s000852/index.shtml

DfE (2010a) The Importance of Teaching: The Schools White Paper 2010 https://www.education.gov.uk/publications/eOrderingDownload/CM-7980.pdf

DfE (2010b) The Importance of Teaching White Paper Equalities Impact Assessment https://www.education.gov.uk/publications/eOrderingDownload/CM-7980-Impact_equalities.pdf

DfE (2011) National Curriculum Assessments at Key Stage 2 in England, 2010/2011 (revised) http://www.education.gov.uk/rsgateway/DB/SFR/s001047/sfr31-2011.pdf

DfE (2012a) Early Years Foundation Stage Profile Attainment by Pupil Characteristics, England 2011/12 http://www.education.gov.uk/rsgateway/DB/SFR/s001098/sfr30-2012.pdf

DfE (2012b) National Curriculum Assessments at Key Stage 2 in England, 2011/2012 (revised) http://www.education.gov.uk/rsgateway/DB/SFR/s001104/sfr33-2012v2.pdf

DfES (2005) *Higher Standards, Better Schools for all: More Choice for Parents and Pupils* https://www.education.gov.uk/publications/eOrderingDownload/Cm%206677.pdf.pdf

Earp (2010) *Unconscious mental processes and the racial achievement gap*. Journal of Multiculturalism in Education, 6:1, p 1-22 http://www.wtamu.edu/webres/File/Journals/MCJ/earp.pdf

Hallam and Ireson (1999) *Pedagogy in the Secondary School.* In Mortimore (Ed) *Understanding Pedagogy*. London: Chapman

Hansen (Ed.) (2012) *Millennium Cohort Study: First, Second, Third and Fourth Surveys*. *A Guide to the Datasets (Seventh Edition).* http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=598&itemtype=document

Hansen and Jones (2011) *Ethnicity and gender gaps in early Childhood.* British Educational Research Journal, 37:6 p 973-991 http://www.tandfonline.com/doi/pdf/10.1080/01411926.2010.515018

Hilton and von Hipple (1996) *Stereotypes.* Annual Review of Psychology 47, p 237–71 http://www.abdn.ac.uk/~psy423/dept/HomePage/Level_3_Social_Psych_files/HiltonVonHippel.pdf

Huang and Gatenby (2010) *Millennium Cohort Study Sweep Four Teacher Survey Technical Report.* http://www.cls.ioe.ac.uk/shared/get-file.ashx?id=489&itemtype=document

Kuklinski and Weinstein (2000) *Classroom and Grade Level Differences in the Stability of Teacher Expectations and Perceived Differential Teacher Treatment*. Learning Environments Research, 3:1, p 1-34

McGarty, Yzerbyt, and Spears (2002) *Stereotypes as Explanations: The Formation of Meaningful Beliefs about Social Groups.* Cambridge University Press

Maylor, Smart, Kuyok and Ross (2009) *Black Children's Achievement Programme Evaluation* https://www.education.gov.uk/publications/eOrderingDownload/DCSF-RR177.pdf

Ofqual (2012) *GCSE English 2012* http://www.ofqual.gov.uk/files/2012-11-02-gcse-english-final-report-and-appendices.pdf

Oxford Brookes University (2005) *Gifted & Talented Co-ordinators' Secondary Professional Development Programme 2005-06 Course Materials: Unit 1, Section 4* http://www.brookes.ac.uk/schools/education/rescon/cpdgifted/docs/unit1/1-4-whatteachersthink.pdf

Reeves, Boyle, and Christie (2001)*The Relationship between Teacher Assessments and Pupil Attainments in Standard Test Tasks at Key Stage 2, 1996-98* British Educational Research Journal 27:2, p 141-160

Strand, Coulon, Meschi, Vorhouse, Frumkin, Ivins, Small, Sood, Gervais, Rehman (2010) *Drivers and Challenges in Raising the Achievement of Pupils from Bangladeshi, Somali and Turkish Backgrounds* https://www.education.gov.uk/publications/eOrderingDownload/DCSF-RR226.pdf

Strand (2011) *The limits of social class in explaining ethnic gaps in educational attainment.* British Educational Research Journal, 37:2, p 197-229

Strand (2012) *The White British–Black Caribbean achievement gap: tests, tiers and teacher expectations* http://www2.warwick.ac.uk/fac/soc/cedar/staff/stevestrand/strand_inpress_tiering_prepub.pdf

Strathern (1997) *'Improving ratings': audit in the British University system.* European Review, 5:3, p 305-321

Thomas, Thomas, Smees, Madaus, and Raczek (1998) *Comparing TeacherAssessment and Standard Task Results in England: the relationship between pupil characteristics and attainment*. Assessment in Education: Principles, Policy & Practice, 5:2, p 213-246 http://www.tandfonline.com/doi/pdf/10.1080/0969594980050205

University of London. Institute of Education. Centre for Longitudinal Studies, Millennium Cohort Study: Fourth Survey, 2008 [computer file]. 3rd Edition. Colchester, Essex: UK Data Archive [distributor], August 2012. SN: 6411, http://dx.doi.org/10.5255/UKDA-SN-6411-2

University of London. Institute of Education. Centre for Longitudinal Studies, Millennium Cohort Study: Fourth Survey, Teacher Survey, 2008 [computer file]. Colchester, Essex: UK Data Archive [distributor], August 2011. SN: 6848 http://dx.doi.org/10.5255/UKDA-SN-6848-1

Wilkin, Derrington, White, Martin, Foster, Kinder, and Rutt (2010) *Improving the outcomes for Gypsy, Roma and Traveller pupils: final report* https://www.education.gov.uk/publications/eOrderingDownload/DFE-RR043.pdf

Annex A: Variables used in main analyses

Measure	Identifier in MCS datasets	Categories before recoding / range	Recoded version
Gender	dhcsexa0	Male / female	-
Ethnicity	ddc08ea0	White / Mixed / Indian / Pakistani / Bangladeshi / Black Caribbean / Black African	-
Language spoken in home	ddhlan00	English only / Mostly English, sometimes other / about half English, half other / mostly other, sometimes English / other language only	English only / Additional languages
SEN	DQ2328	Yes / No	-
Income-level	doedp000	Above 60% median / Below 60% median	-
BAS Word Reading (Ability) Score (a)	dcwrab00	10 - 214	-
BAS Word Reading (Ability) Score (b) (quintiles)	dcwrab00	10 - 214	Bottom quintile (10 - 85) / Second quintile (66 - 102) / Middle quintile (103 – 116) / Fourth quintile (117 – 133) / Top quintile (134 – 214)
Progress in Maths score (a)	maths7scale	0 - 28	-
Progress in Maths score (b) (quintiles)	maths7scale	0 - 28	Bottom quintile (0 – 12.7) / Second quintile (12.8 - 17) / Middle quintile (18 – 20) / Fourth quintile (21 – 23) / Top quintile (24 – 28)
GOR (a)	DDREGN00	North East / North West / Yorks and Humber / East Midlands / West Midlands / East England / London / South East / South West	-
GOR (b) (merged, seven areas - EAL models)	DDREGN00	North East / North West / Yorks and Humber / East Midlands / West Midlands / East England / London / South East / South West	North East and Yorks and Humber / North West / East Midlands / West Midlands / East England / London / South East and South West
GOR (c) (merged, four areas - ethnicity	DDREGN00	North East / North West / Yorks and Humber / East Midlands / West Midlands / East England / London / South East / South West	North East, North West and Yorks and Humber / East Midlands and

models)			West Midlands / East England, South East and South West / London
Sample strata	ptty00	Advantaged / Disadvantaged / Ethnic	-
Main parent qualification level	DMDNVQ00	Level 1 / Level 2 / Level 3 / Level 4 / Level 5 / Overseas qualification / None	-
Main parent NS-SEC	DMD05C00	Managerial and professional / Intermediate / Small employer and self-employed / Lower supervisory and technical / Semi-routine and routine	-
Teacher judgement of reading ability	DQ2162	Well above average / Above average / Average / Below average / Well below average	Above average / Not above average; Below average / Not below average
Teacher judgement of maths ability	DQ2167	Well above average / Above average / Average / Below average / Well below average	Above average / Not above average; Below average / Not below average

Annex B: Testing whether differences in timings of pupil test and teacher survey completion may affect results

Fieldwork for the main MCS wave four survey and for the subsequent teacher survey took place across a number of sequential waves (Chaplin Grey et al, 2010; Huang and Gatenby, 2010). Given, particularly, the uneven distribution of pupils across areas according to some relevant characteristics (see Tables 13-14, Annex D), it is possible that there might be differences in pupil age at test completion and / or pupil age at teacher survey completion – and also, therefore, in intervening time lag. Any variation may correspond to pupil characteristics such as ethnicity or income, and might affect results.

To test whether this is the case, initial analysis according to each characteristic, for each teacher judgement (as reported in Figures 8-11) is repeated, incorporating controls for:

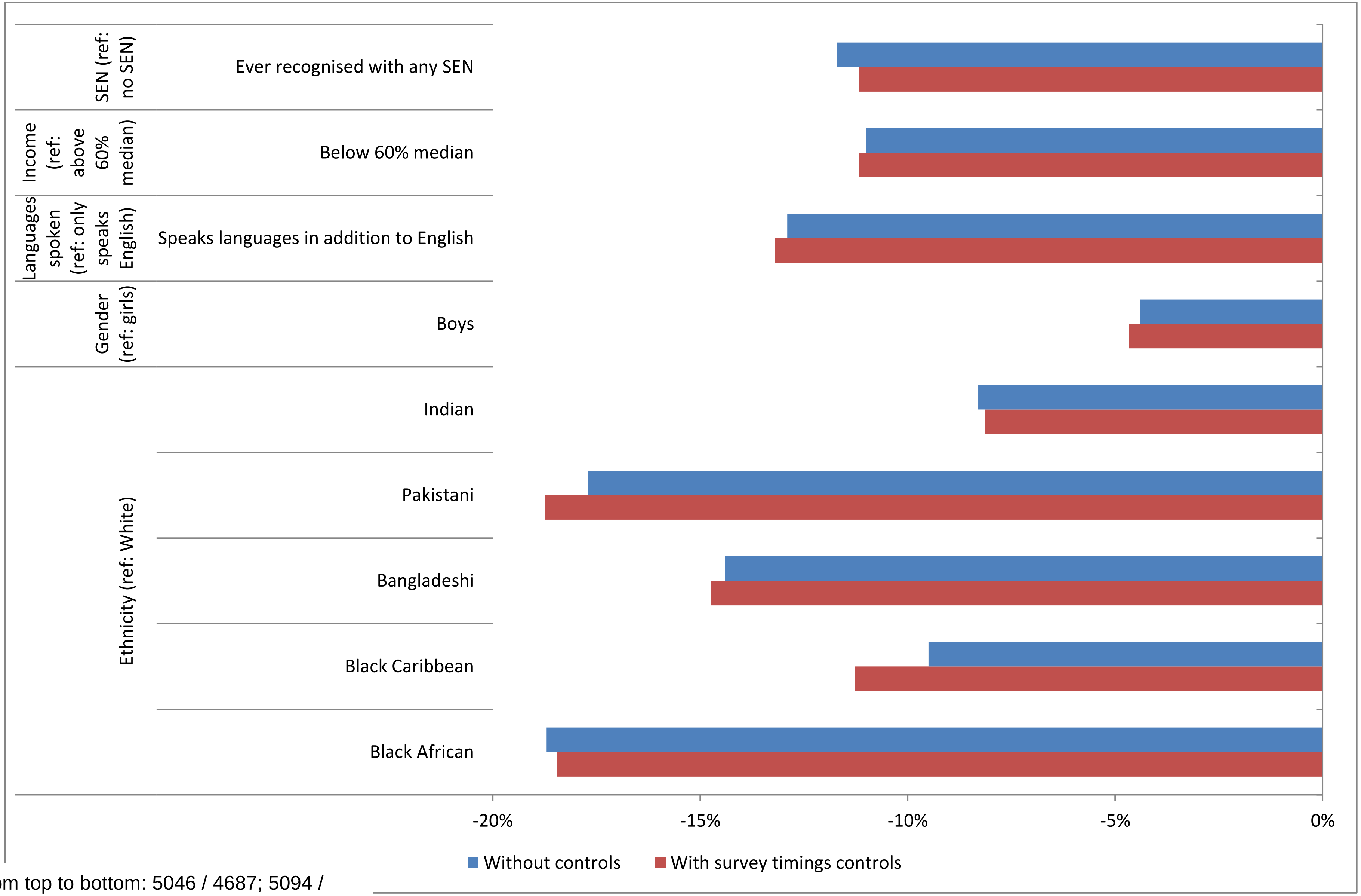
- pupil age in months at cognitive test completion;[27]
- time elapsed in months between cognitive test completion and teacher survey completion.[28]

Little difference is made to the magnitude and none to the direction of findings – as illustrated in Figures 20 to 23. Any systematic differences by pupil group in age at completing cognitive tests and time lag between test completion and teacher survey do not account for biases according to pupil characteristics.

[27] Using variable *age*.

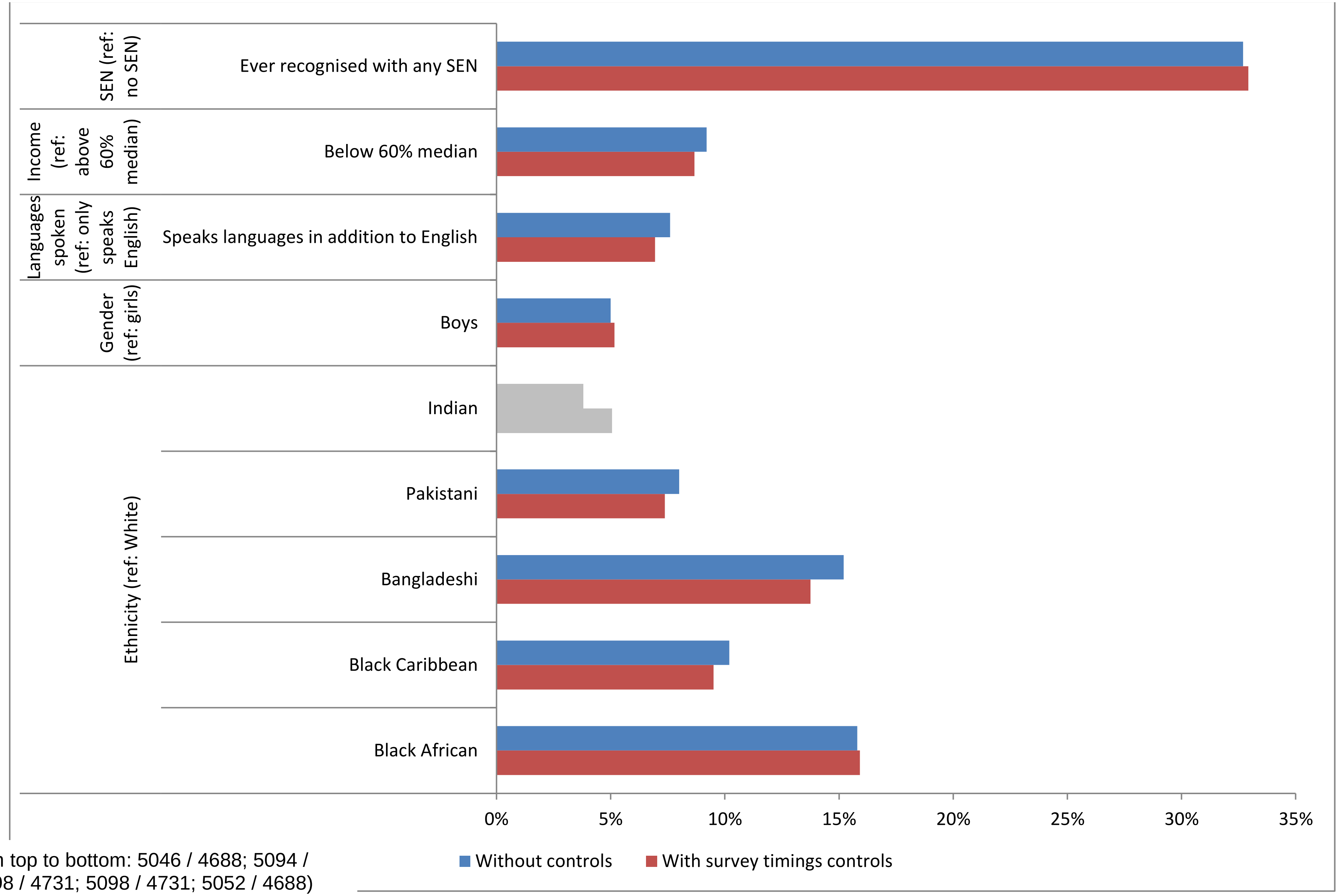
[28] This is constructed using variables *DQ2702, DQ2704*, *dhcdbma0, age.* Firstly, pupil age in months at teacher survey is calculated using information on month and year of birth (*dhcdbma0*; all English cohort children were born over one academic year, between September 2000 and August 2001 [Hansen [ed], 2012]) and month and year of teacher survey completion (*DQ2702, DQ2704*). Extremes are removed from the latter; only teacher responses recorded as taking place within the main survey window (June 2008-May 2009) are included. Then age in months at cognitive test completion (*age)* is subtracted from age in months at teacher survey, giving time in months between pupil cognitive test and teacher survey completion.

Figure 20: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'above average' at reading by their teacher, compared to pupils with the reference characteristic – with timing controls



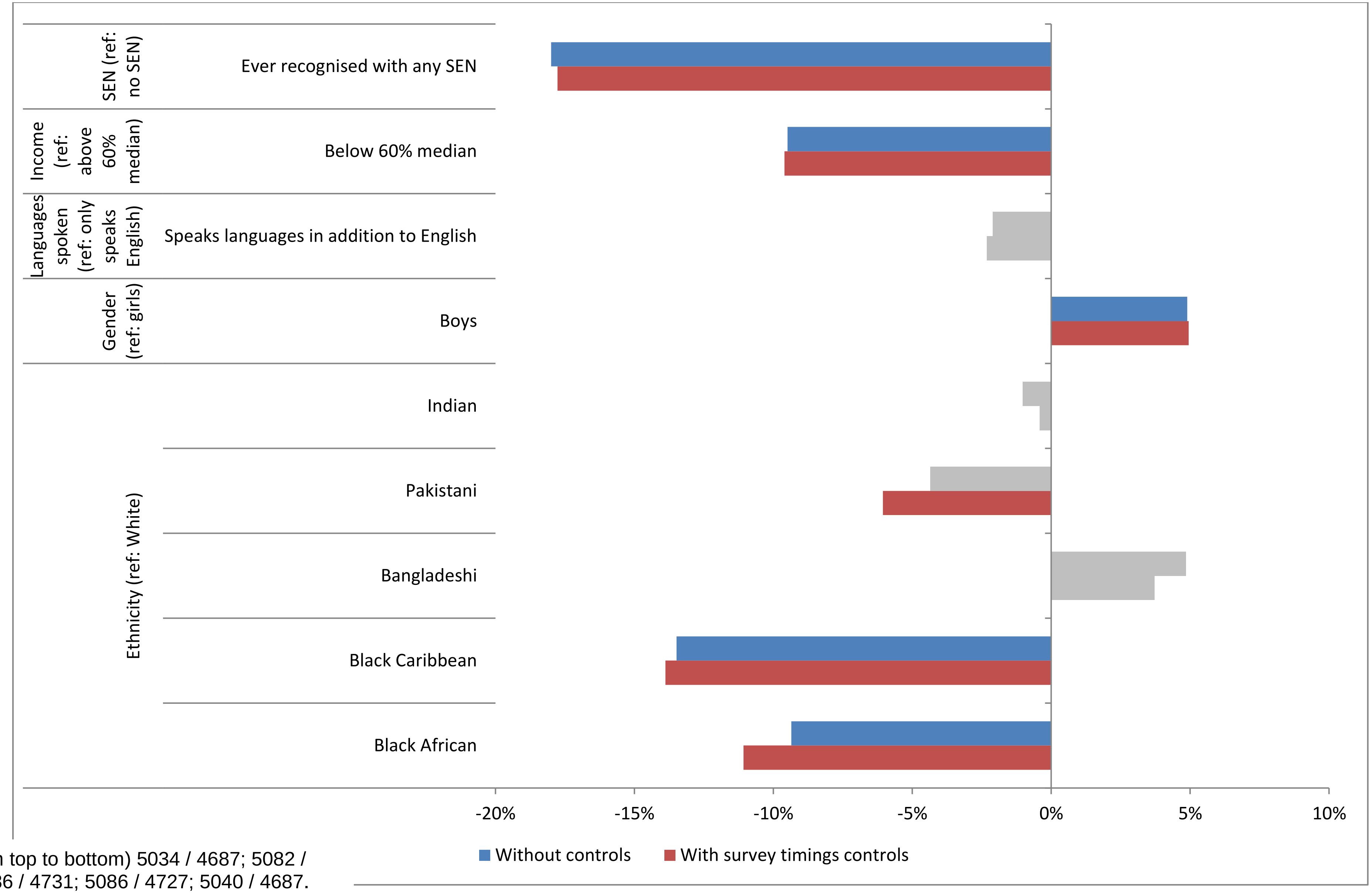
ns = (from top to bottom: 5046 / 4687; 5094 / 4727; 5098 / 4731; 5098 / 4731; 5052 / 4688)

Figure 21: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'below average' at reading by their teacher, compared to pupils with the reference characteristic – with timing controls



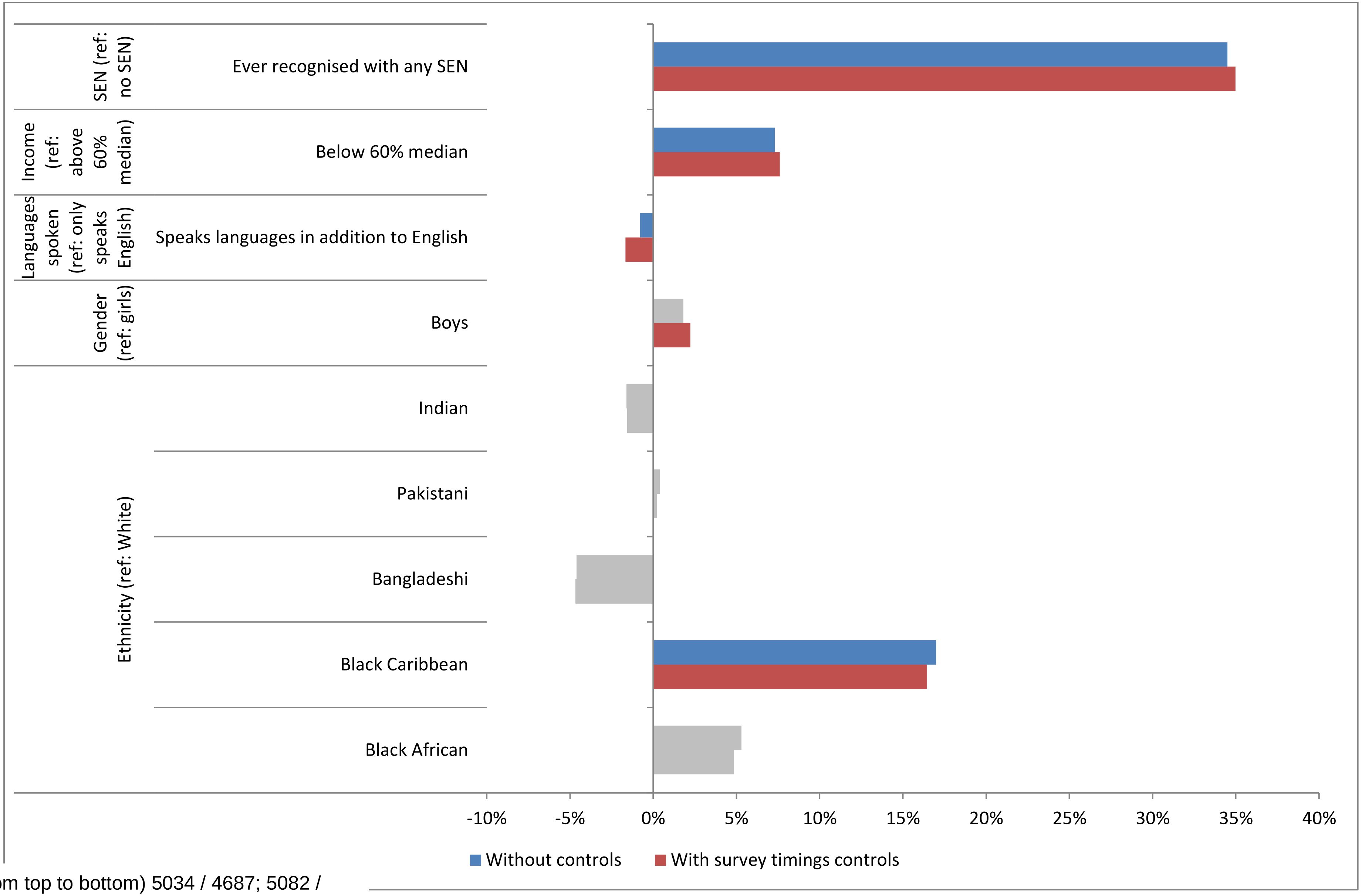
ns = (from top to bottom: 5046 / 4688; 5094 / 4727; 5098 / 4731; 5098 / 4731; 5052 / 4688)
Grey bars not significant at the 5% level

Figure 22: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'above average' at maths by their teacher, compared to pupils with the reference characteristic – with timing controls



ns = (from top to bottom) 5034 / 4687; 5082 / 4731; 5086 / 4731; 5086 / 4727; 5040 / 4687.
Grey bars not significant at the 5% level

Figure 23: Difference in percentage point likelihood of pupils with each respective characteristic being judged 'below average' at maths by their teacher, compared to pupils with the reference characteristic – with timing controls

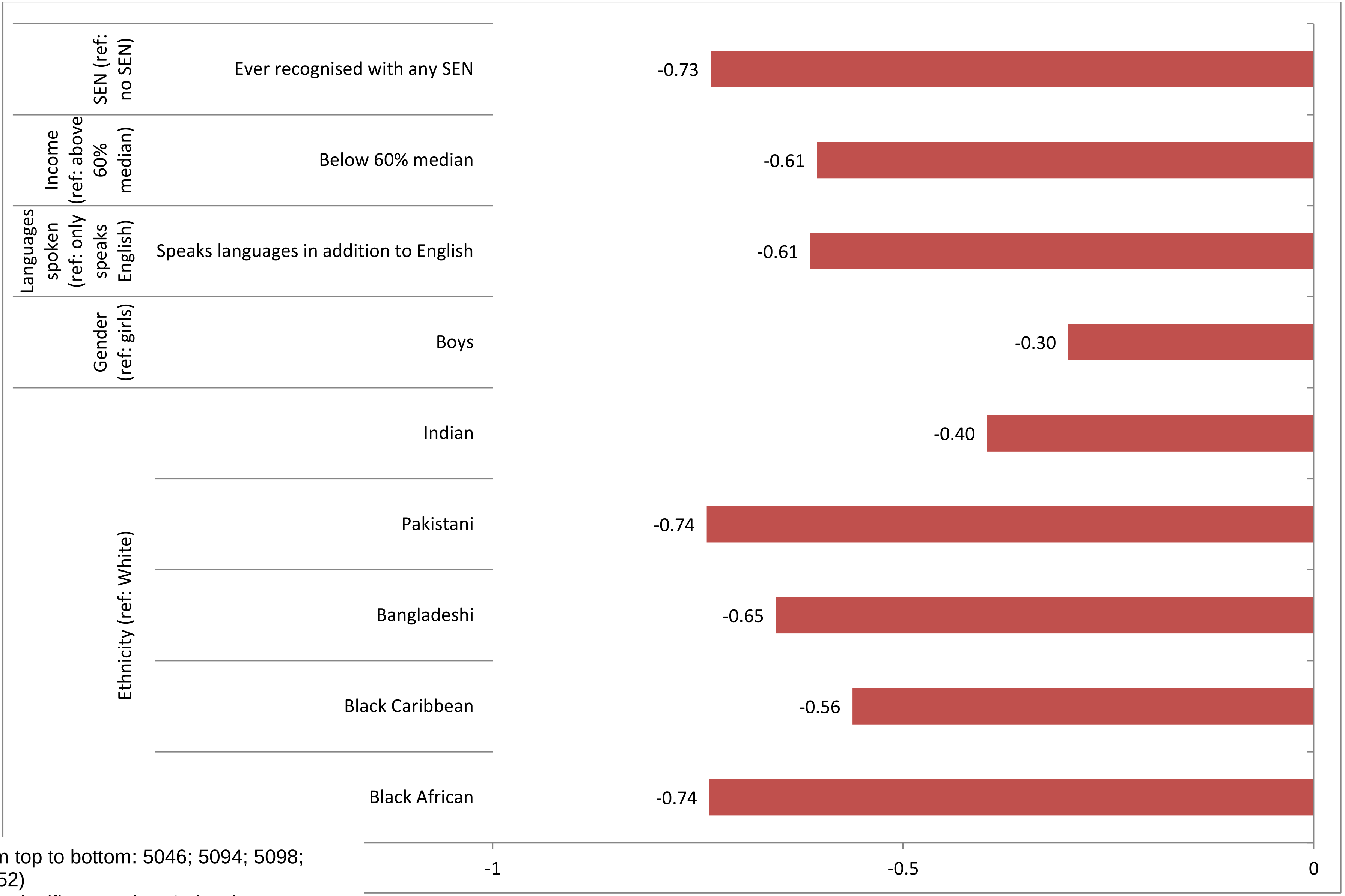


ns = (from top to bottom) 5034 / 4687; 5082 / 4727; 5086 / 4731; 5086 / 4731; 5040 / 4687. Grey bars not significant at the 5% level

Annex C: Binary logistic versions of models

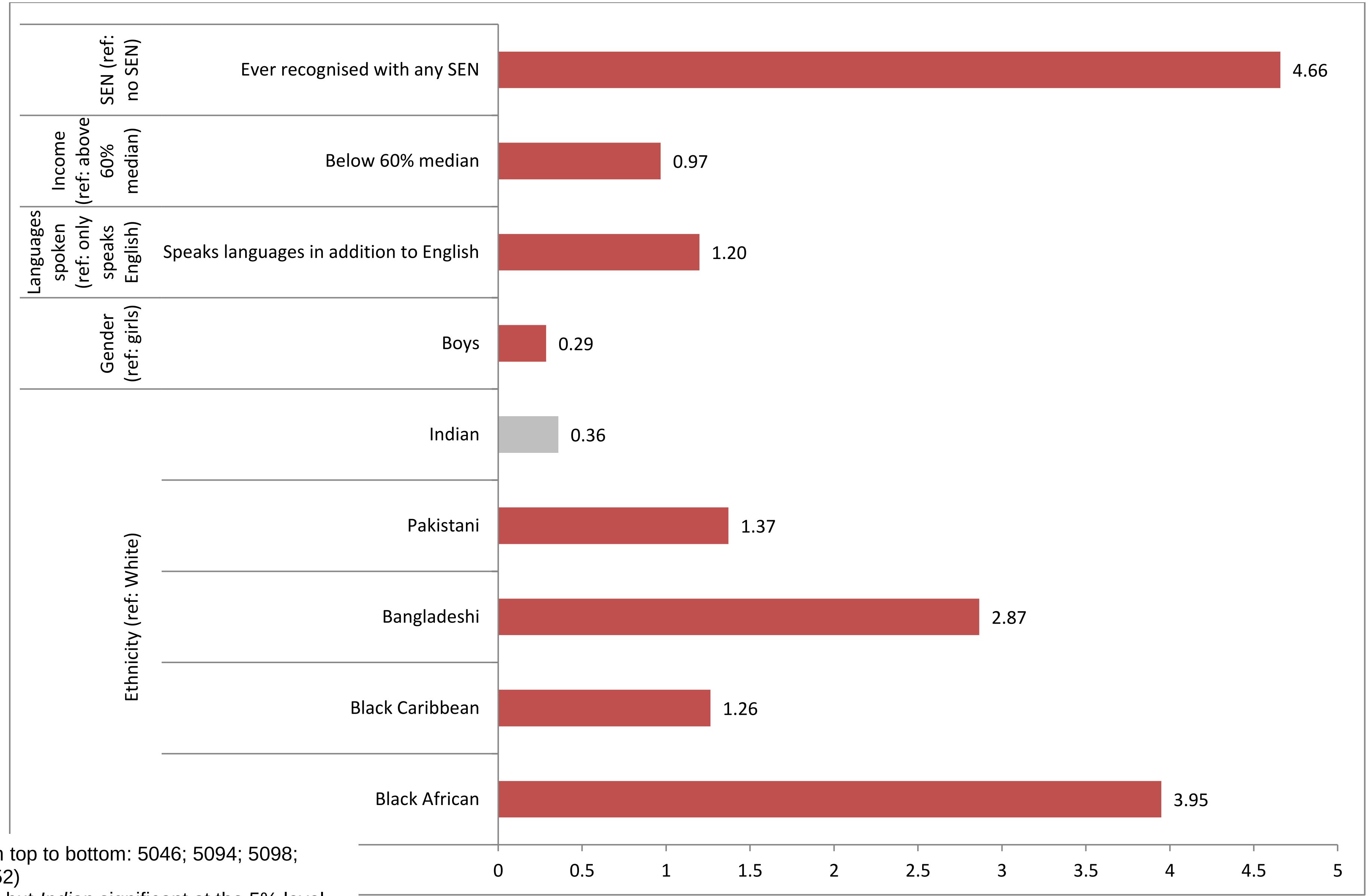
In order to check that choice of a linear probability model has not influenced the direction of results, initial analysis according to each characteristic, for each teacher judgement (as reported in Figures 8-11) is repeated using a binary logistic model. Figures 24 to 27, below, report the difference in odds for pupils with each characteristic (from each respective reference category, whose value is 1, using the exponentiated coefficient). No difference is made by use of the linear model to the direction of results.

Figure 24: Binary logistic model – difference in odds of pupils with each respective characteristic being judged 'above average' at reading by their teacher, compared to pupils with the reference characteristic



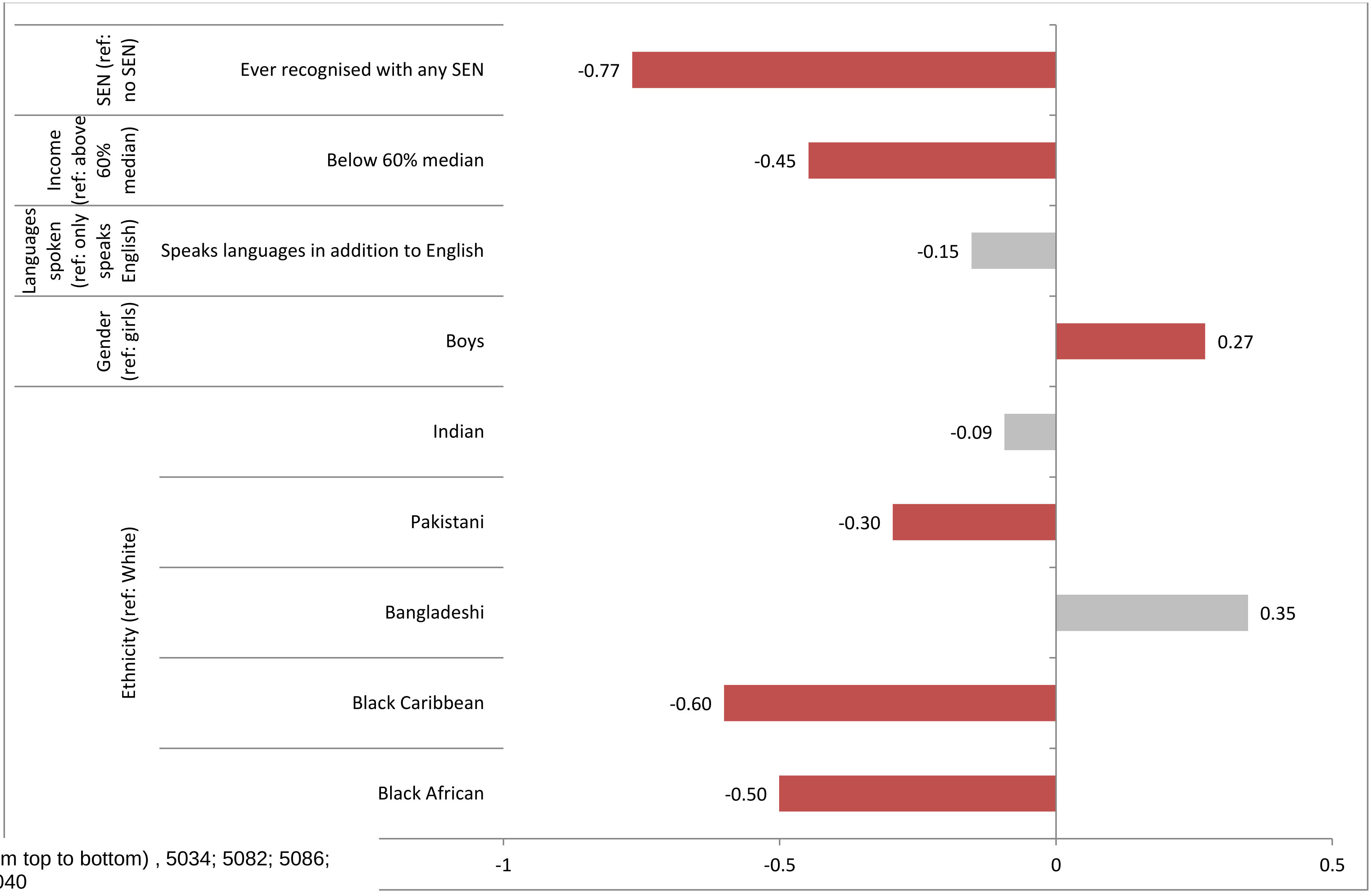
ns = (from top to bottom: 5046; 5094; 5098; 5098; 5052)
All results significant at the 5% level

Figure 25: Binary logistic model – difference in odds of pupils with each respective characteristic being judged 'below average' at reading by their teacher, compared to pupils with the reference characteristic



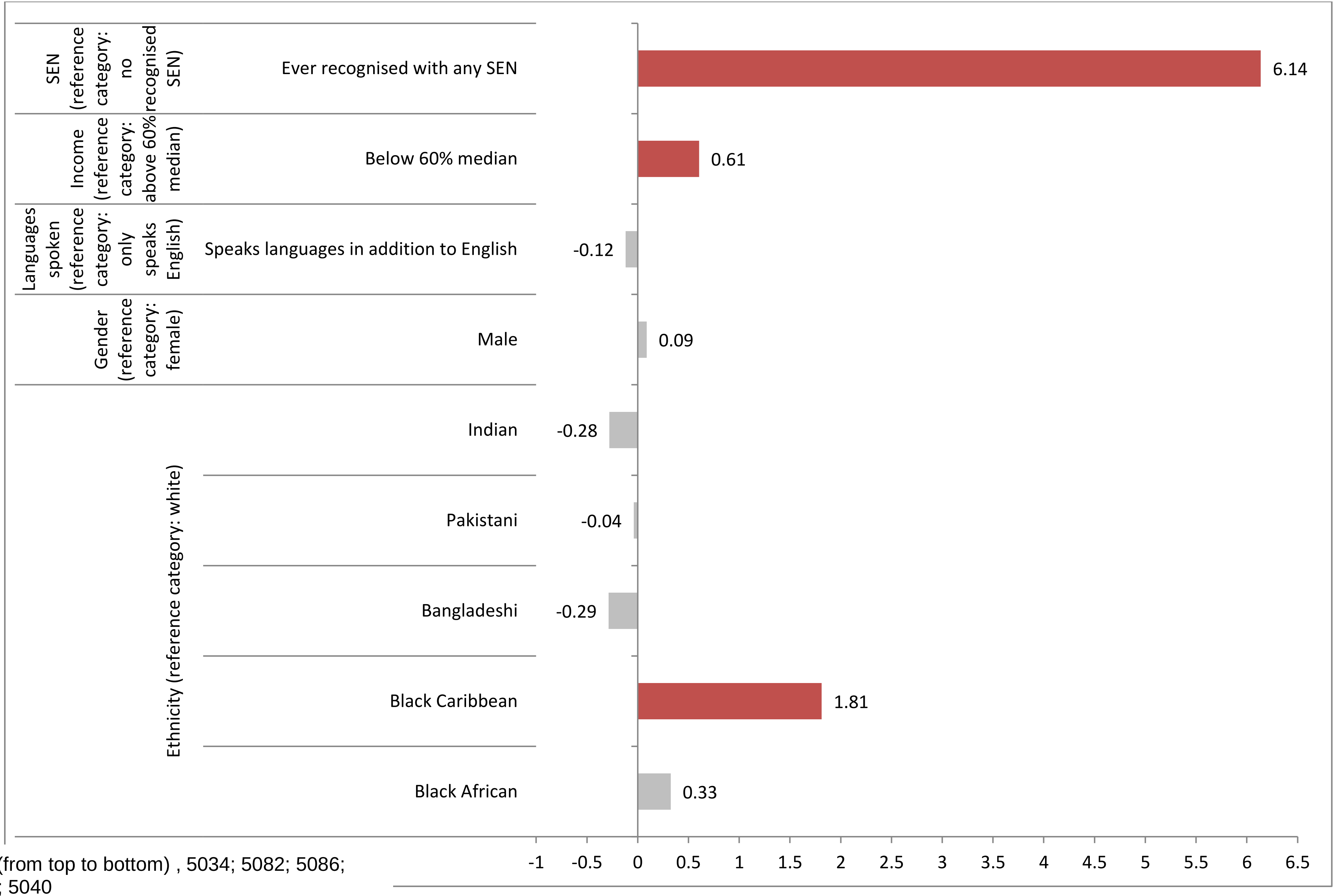
ns = (from top to bottom: 5046; 5094; 5098; 5098; 5052)
All results but *Indian* significant at the 5% level

Figure 26: Binary logistic model – difference in odds of pupils with each respective characteristic being judged 'above average' at maths by their teacher, compared to pupils with the reference characteristic



ns = (from top to bottom) , 5034; 5082; 5086; 5086; 5040
Red bars significant at the 5% level

Figure 27: Binary logistic model – difference in odds of pupils with each respective characteristic being judged ‘below average’ at maths by their teacher, compared to pupils with the reference characteristic



ns = (from top to bottom) , 5034; 5082; 5086; 5086; 5040
Red bars significant at the 5% level

Annex D: Model Coefficients

Each set of tables corresponds to a set of results presented in the main text, as referenced. In all tables, standard errors are presented in brackets, and:

- *** = p < .001;
- ** = p < .05;
- *= p < .10.

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at reading (see Figure 8)

SEN model (n = 5046)	
SEN (ref = no SEN)	-.117 (.013)***
Word Reading score	.010 (.000)***
Intercept	-.695 (.021)***

Income model (n = 5094)	
Low income (ref = higher income)	-.110 (.012)***
Word Reading score	.010 (.000)***
Intercept	-.623 (.022)***

Language model (n = 5098)	
Other languages (ref = English only)	-.129 (.016)***
Word Reading score	0.11 (.000)***
Intercept	-.809 (.025)***

Gender model (n = 5098)	
Boy (ref = girl)	-.044 (.011)***
Word Reading score	.010 (.000)***
Intercept	-.707 (.021)***

Ethnicity model (n = 5052)	
Indian (ref = White)	-.083 (.032)**
Pakistani (ref = White)	-.177 (.024)***
Bangladeshi (ref = White)	-.144 (.042)**
Black Caribbean (ref= White)	-.095 (.047)**
Black African (ref = White)	-.187 (.037)***
Word Reading score	.011 (.000)***
Intercept	-.677 (.021)***

Biases in teacher judgements of whether pupils are of 'below average' ability and attainment at reading (see Figure 9)

SEN model (n = 5046)	
SEN (ref = no SEN)	.327 (.016)***
Word Reading score	-.006 (.000)***
Intercept	1.174 (.016)***

Income model (n = 5094)	
Low income (ref = higher income)	.092 (.010)***
Word Reading score	-.008 (.000)***
Intercept	1.103 (.018)***

Language model (n = 5098)	
Other languages (ref = English only)	.074 (.013)***
Word Reading score	-.009 (.000)***
Intercept	1.228 (.020)***

Ethnicity model (n = 5052)	
Indian (ref = White)	.038 (.026)
Pakistani (ref = White)	.080 (.020)***
Bangladeshi (ref = White)	.152 (.035)***
Black Caribbean (ref= White)	.102 (.039)**
Black African (ref = White)	.159 (.030)***
Word Reading score	-.009 (.000)***
Intercept	1.150 (.017)***

Gender model (n=5098)	
Boy (ref = girl)	.050 (.009)***
Word Reading score	-.009 (.000)***
Intercept	1.178 (.017)***

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at maths (See Figure 10)

SEN model (n = 5034)	
SEN (ref = no SEN)	-.180 (.015)***
Progress in Maths score	.037 (.001)***
Intercept	-.413 (.020)***

Income model (n = 5082)	
Low income (ref = higher income)	-.095 (.014)***
Progress in Maths score	.040 (.001)***
Intercept	-.308 (.022)***

Language model (n = 5086)	
Other languages (ref = English only)	-.021 (.018)
Progress in Maths score	.041 (.001)***
Intercept	-.381 (.024)***

Gender model (n = 5086)	
Boy (ref = girl)	.049 (.012)***
Progress in Maths score	.042 (.001)***
Intercept	-.341 (.021)***

Ethnicity model (n = 5040)	
Indian (ref = White)	-.010 (.035)
Pakistani (ref = White)	-.043 (.027)
Bangladeshi (ref = White)	.049 (.046)
Black Caribbean (ref= White)	-.135 (.052)**
Black African (ref = White)	-.094 (.041)**
Progress in Maths score	.041 (.001)***
Intercept	-.357 (.021)***

Biases in teacher judgements of whether pupils are of 'below average' ability and attainment at maths (see Figure 11)

SEN model (n = 5034)	
SEN (ref = no SEN)	.345 (.012)***
Progress in Maths score	-.024 (.000)***
Intercept	.912 (.016)***

Income model (n = 5082)	
Low income (ref = higher income)	.073 (.011)***
Progress in Maths score	-.032 (.000)***
Intercept	.776 (.018)***

Language model (n = 5086)	
Other languages (ref = English only)	-.008 (.014)
Progress in Maths score	-.033 (.001)***
Intercept	.815 (.020)***

Gender model (n = 5086)	
Boy (ref = girl)	.018 (.010)*
Progress in Maths score	-.033 (.000)***
Intercept	.830 (.017)***

Ethnicity model (n = 5040)	
Indian (ref = White)	-.016 (.028)
Pakistani (ref = White)	.004 (.022)
Bangladeshi (ref = White)	-.046 (.038)
Black Caribbean (ref= White)	.170 (.043)***
Black African (ref = White)	.053 (.034)
Progress in Maths score	-.033 (.000)***
Intercept	.817 (.017)***

Differences in bias according to cognitive test score level: reading 'above average' (see Table 6)

SEN model (n = 5046)	
SEN (ref = no SEN)	-.307 (.035)***
Bottom quintile (ref = middle quintile)	-.443 (.023)***

Second lowest quintile (ref = middle quintile)	-.248 (.019)***
Second highest quintile (ref = middle quintile)	.260 (.017)***
Top quintile (ref = middle quintile)	.431 (.017)***
SEN x bottom quintile	.278 (.043)***
SEN x second lowest quintile	.139 (.045)**
SEN x second highest quintile	.048 (.056)
SEN x top quintile	.223 (.057)***
Intercept	.472 (.013)***

Income model (n = 5094)	
Low income (ref = higher income)	-.213 (.027)***
Bottom quintile (ref = middle quintile)	-.463 (.020)***
Second lowest quintile (ref = middle quintile)	-.279 (.020)***
Second highest quintile (ref = middle quintile)	.274 (.019)***
Top quintile (ref = middle quintile)	.443 (.018)***
Low income x bottom quintile	.197 (.040)***
Low income x second lowest quintile	.147 (.038)***
Low income x second highest quintile	.041 (.039)
Low income x top quintile	.079 (.039)**
Intercept	.481 (.014)***

Language model (n = 5098)	
Other languages (ref = English only)	-.200 (.034)***
Bottom quintile (ref = middle quintile)	-.443 (.018)***
Second lowest quintile (ref = middle quintile)	-.257 (.018)***
Second highest quintile (ref = middle quintile)	.277 (.018)***
Top quintile (ref = middle quintile)	.461 (.018)***
Other languages x bottom quintile	.197 (.050)***
Other languages x second lowest quintile	.088 (.050)*
Other languages x second highest quintile	.058 (.050)
Other languages x top quintile	.084 (.048)*
Intercept	.455 (.013)***

Gender model (n = 5086)	
Boy (ref = girl)	-.054 (.024)**
Bottom quintile (ref = middle quintile)	-.439 (.025)***
Second lowest quintile (ref = middle quintile)	-.240 (.024)***
Second highest quintile (ref = middle quintile)	.290 (.022)***
Top quintile (ref = middle quintile)	.455 (.023)***
Boys x bottom quintile	.047 (.034)
Boys x second lowest quintile	-.009 (.034)
Boys x second highest quintile	-.022 (.033)
Boys x top quintile	.031 (.033)
Intercept	.454 (.016)***

Ethnicity model (n = 5052)	
Indian (ref = White)	-.165 (.069)**
Pakistani (ref = White)	-.229 (.051)***
Bangladeshi (ref = White)	-.298 (.107)**
Black Caribbean (ref = White)	-.215 (.107)**
Black African (ref = White)	-.406 (.090)***
Bottom quintile (ref = middle quintile)	-.453 (.018)***
Second lowest quintile (ref = middle quintile)	-.263 (.019)***

Second highest quintile (ref = middle quintile)	.272 (.018)***
Top quintile (ref = middle quintile)	.451 (.018)***
Indian x bottom quintile	.201 (.106)*
Indian x second lowest quintile	.088 (.103)
Indian x second highest quintile	.046 (.094)
Indian x top quintile	.180 (.090)**
Pakistani x bottom quintile	.217 (.071)**
Pakistani x second lowest quintile	.062 (.072)
Pakistani x second highest quintile	-.043 (.072)
Pakistani x top quintile	.106 (.073)
Bangladeshi x bottom quintile	.286 (.149)*
Bangladeshi x second lowest quintile	.263 (.132)**
Bangladeshi x second highest quintile	.362 (.159)**
Bangladeshi x top quintile	.086 (.129)
Black Caribbean x bottom quintile	.203 (.135)
Black Caribbean x second lowest quintile	.104 (.155)
Black Caribbean x second highest quintile	.078 (.159)
Black Caribbean x top quintile	.157 (.146)
Black African x bottom quintile	.394 (.126)**
Black African x second lowest quintile	.340 (.120)**
Black African x second highest quintile	.246 (.116)**
Black African x top quintile	.224 (.113)**
Intercept	.465 (.013)***

Differences in bias according to cognitive test score level: reading ‘below average’ (see Table 7)

SEN model (n = 5046)	
SEN (ref = no SEN)	.192 (.026)***
Bottom quintile (ref = middle quintile)	.469 (.017)***
Second lowest quintile (ref = middle quintile)	.120 (.014)***
Second highest quintile (ref = middle quintile)	-.036 (.013)**
Top quintile (ref = middle quintile)	-.042 (.013)**
SEN x bottom quintile	.192 (.032)***
SEN x second lowest quintile	.157 (.034)***
SEN x second highest quintile	-.095 (.042)**
SEN x top quintile	-.181 (.043)**
Intercept	.045 (.009)***

Income model (n = 5094)	
Low income (ref = higher income)	.083 (.022)***
Bottom quintile (ref = middle quintile)	.633 (.017)***
Second lowest quintile (ref = middle quintile)	.173 (.016)***
Second highest quintile (ref = middle quintile)	-.041 (.015)**
Top quintile (ref = middle quintile)	-.048 (.015)**
Low income x bottom quintile	.071 (.029)**
Low income x second lowest quintile	.015 (.030)
Low income x second highest quintile	-.057 (.031)*
Low income x top quintile	-.076 (.032)**
Intercept	.051 (.012)***

Language model (n = 5098)	
Other languages (ref = English only)	.069 (.029)**
Bottom quintile (ref = middle quintile)	.682 (.014)***
Second lowest quintile (ref = middle quintile)	.169 (.015)***
Second highest quintile (ref = middle quintile)	-.048 (.014)**
Top quintile (ref = middle quintile)	-.059 (.014)***
Other languages x bottom quintile	.006 (.040)
Other languages x second lowest quintile	.093 (.040)**
Other languages x second highest quintile	-.062 (.040)
Other languages x top quintile	-.066 (.038)*
Intercept	.062 (.010)***

Gender model (n = 5086)	
Boy (ref = girl)	-.002 (.019)
Bottom quintile (ref = middle quintile)	.618 (.020)***
Second lowest quintile (ref = middle quintile)	.158 (.019)***
Second highest quintile (ref = middle quintile)	-.058 (.018)**
Top quintile (ref = middle quintile)	-.070 (.018)***
Boys x bottom quintile	.108 (.027)***
Boys x second lowest quintile	.050 (.027)*
Boys x second highest quintile	.007 (.027)
Boys x top quintile	.005 (.026)
Intercept	.071 (.013)***

Ethnicity model (n = 5052)	
Indian (ref = White)	.045 (.055)
Pakistani (ref = White)	.090 (.041)**
Bangladeshi (ref = White)	.278 (.086)**
Black Caribbean (ref = White)	.111 (.086)
Black African (ref = White)	.180 (.072)**
Bottom quintile (ref = middle quintile)	.688 (.015)***
Second lowest quintile (ref = middle quintile)	.167 (.015)***
Second highest quintile (ref = middle quintile)	-.041 (.014)**
Top quintile (ref = middle quintile)	-.053 (.015)***
Indian x bottom quintile	-.121 (.085)
Indian x second lowest quintile	.025 (.082)
Indian x second highest quintile	-.029 (.075)
Indian x top quintile	-.047 (.072)
Pakistani x bottom quintile	.014 (.057)
Pakistani x second lowest quintile	.045 (.058)
Pakistani x second highest quintile	-.070 (.057)
Pakistani x top quintile	-.093 (.059)
Bangladeshi x bottom quintile	-.098 (.119)
Bangladeshi x second lowest quintile	-.084 (.105)
Bangladeshi x second highest quintile	-.294 (.127)**
Bangladeshi x top quintile	-.243 (.103)**
Black Caribbean x bottom quintile	-.093 (.108)
Black Caribbean x second lowest quintile	.212 (.124)*
Black Caribbean x second highest quintile	-.126 (.127)
Black Caribbean x top quintile	-.114 (.117)
Black African x bottom quintile	.021 (.101)
Black African x second lowest quintile	.143 (.094)

Black African x second highest quintile	-.156 (.093)*
Black African x top quintile	-.182 (.091)**
Intercept	.055 (.010)***

Differences in bias according to cognitive test score level: maths 'above average' (see Table 8)

SEN model (n = 5034)	
SEN (ref = no SEN)	-.271 (.039)***
Bottom quintile (ref = middle quintile)	-.344 (.024)***
Second lowest quintile (ref = middle quintile)	-.191 (.022)***
Second highest quintile (ref = middle quintile)	.182 (.020)***
Top quintile (ref = middle quintile)	.306 (.020)***
SEN x bottom quintile	.175 (.047)***
SEN x second lowest quintile	.109 (.050)**
SEN x second highest quintile	-.091 (.055)*
SEN x top quintile	-.006 (.060)
Intercept	.446 (.015)***

Income model (n = 5082)	
Low income (ref = higher income)	-.186 (.033)***
Bottom quintile (ref = middle quintile)	-.379 (.024)***
Second lowest quintile (ref = middle quintile)	-.215 (.023)***
Second highest quintile (ref = middle quintile)	.161 (.021)***
Top quintile (ref = middle quintile)	.314 (.022)***
Low income x bottom quintile	.155 (.043)***
Low income x second lowest quintile	.132 (.044)**
Low income x second highest quintile	.070 (.045)
Low income x top quintile	.009 (.047)
Intercept	.444 (.016)***

Language model (n = 5086)	
Other languages (ref = English only)	-.063 (.048)
Bottom quintile (ref = middle quintile)	-.359 (.021)***
Second lowest quintile (ref = middle quintile)	-.197 (.021)***
Second highest quintile (ref = middle quintile)	.179 (.020)***
Top quintile (ref = middle quintile)	.326 (.020)***
Other languages x bottom quintile	.074 (.058)
Other languages x second lowest quintile	.064 (.062)
Other languages x second highest quintile	.015 (.061)
Other languages x top quintile	-.012 (.063)
Intercept	.408 (.015)***

Gender model (n = 5086)	
Boy (ref = girl)	.046 (.028)*
Bottom quintile (ref = middle quintile)	-.316 (.027)***
Second lowest quintile (ref = middle quintile)	-.201 (.027)***
Second highest quintile (ref = middle quintile)	.181 (.026)***
Top quintile (ref = middle quintile)	.292 (.027)***
Boys x bottom quintile	-.071 (.039)*
Boys x second lowest quintile	.021 (.039)
Boys x second highest quintile	-.005 (.038)
Boys x top quintile	.051 (.039)
Intercept	.380 (.019)***

Ethnicity model (n = 5040)	
Indian (ref = White)	.018 (.107)
Pakistani (ref = White)	-.194 (.077)**
Bangladeshi (ref = White)	-.056 (.128)
Black Caribbean (ref = White)	-.320 (.135)**
Black African (ref = White)	-.170 (.107)
Bottom quintile (ref = middle quintile)	-.371 (.022)***
Second lowest quintile (ref = middle quintile)	-.208 (.022)***
Second highest quintile (ref = middle quintile)	.163 (.021)***
Top quintile (ref = middle quintile)	.314 (.021)***
Indian x bottom quintile	-.023 (.139)
Indian x second lowest quintile	.020 (.131)
Indian x second highest quintile	-.033 (.132)
Indian x top quintile	-.045 (.123)
Pakistani x bottom quintile	.182 (.089)**
Pakistani x second lowest quintile	.157 (.095)
Pakistani x second highest quintile	.102 (.098)
Pakistani x top quintile	.211 (.112)*
Bangladeshi x bottom quintile	.096 (.148)
Bangladeshi x second lowest quintile	.138 (.165)
Bangladeshi x second highest quintile	.105 (.161)
Bangladeshi x top quintile	.156 (.215)
Black Caribbean x bottom quintile	.217 (.164)*
Black Caribbean x second lowest quintile	.108 (.197)
Black Caribbean x second highest quintile	.283 (.186)
Black Caribbean x top quintile	.202 (.179)
Black African x bottom quintile	.151 (.131)
Black African x second lowest quintile	.180 (.147)
Black African x second highest quintile	.153 (.139)
Black African x top quintile	-.216 (.139)
Intercept	.420 (.015)***

Differences in bias according to cognitive test score level: maths 'below average' (see Table 9)

SEN model (n = 5034)	
SEN (ref = no SEN)	.274 (.030)***
Bottom quintile (ref = middle quintile)	.297 (.019)***
Second lowest quintile (ref = middle quintile)	.078 (.017)***
Second highest quintile (ref = middle quintile)	-.041 (.015)**
Top quintile (ref = middle quintile)	-.055 (.016)***
SEN x bottom quintile	.109 (.036)**
SEN x second lowest quintile	.144 (.038)***
SEN x second highest quintile	.011 (.043)
SEN x top quintile	-.044 (.046)
Intercept	.067 (.012)***

Income model (n = 5082)	
Low income (ref = higher income)	.124 (.028)***
Bottom quintile (ref = middle quintile)	.457 (.020)***
Second lowest quintile (ref = middle quintile)	.144 (.019)***
Second highest quintile (ref = middle quintile)	-.031 (.018)*
Top quintile (ref = middle quintile)	-.055 (.018)**
Low income x bottom quintile	-.064 (.052)*

Low income x second lowest quintile	-.034 (.036)
Low income x second highest quintile	-.064 (.037)*
Low income x top quintile	-.088 (.039)**
Intercept	.082 (.013)***

Language model (n = 5086)	
Other languages (ref = English only)	.030 (.040)
Bottom quintile (ref = middle quintile)	.473 (.017)***
Second lowest quintile (ref = middle quintile)	.149 (.017)***
Second highest quintile (ref = middle quintile)	-.046 (.016)**
Top quintile (ref = middle quintile)	-.076 (.017)***
Other languages x bottom quintile	-.100 (.048)**
Other languages x second lowest quintile	-.039 (.051)
Other languages x second highest quintile	-.021 (.050)
Other languages x top quintile	-.019 (.052)
Intercept	.108 (.012)***

Gender model (n = 5086)	
Boy (ref = girl)	.003 (.023)
Bottom quintile (ref = middle quintile)	.421 (.023)***
Second lowest quintile (ref = middle quintile)	.149 (.022)***
Second highest quintile (ref = middle quintile)	-.051 (.022)**
Top quintile (ref = middle quintile)	-.073 (.023)**
Boys x bottom quintile	.065 (.032)**
Boys x second lowest quintile	-.009 (.032)
Boys x second highest quintile	.006 (.031)
Boys x top quintile	-.008 (.032)
Intercept	.110 (.016)***

Ethnicity model (n = 5040)	
Indian (ref = White)	-.041 (.088)
Pakistani (ref = White)	.090 (.064)
Bangladeshi (ref = White)	-.012 (.106)
Black Caribbean (ref = White)	.197 (.111)*
Black African (ref = White)	.022 (.088)
Bottom quintile (ref = middle quintile)	.464 (.018)***
Second lowest quintile (ref = middle quintile)	.157 (.018)***
Second highest quintile (ref = middle quintile)	-.041 (.017)**
Top quintile (ref = middle quintile)	-.070 (.017)***
Indian x bottom quintile	-.005 (.115)
Indian x second lowest quintile	-.032 (.109)
Indian x second highest quintile	-.021 (.110)
Indian x top quintile	.047 (.102)
Pakistani x bottom quintile	-.139 (.074)*
Pakistani x second lowest quintile	-.049 (.079)
Pakistani x second highest quintile	-.074 (.081)
Pakistani x top quintile	-.123 (.093)
Bangladeshi x bottom quintile	-.055 (.123)
Bangladeshi x second lowest quintile	-.012 (.137)
Bangladeshi x second highest quintile	-.050 (.134)
Bangladeshi x top quintile	-.021 (.178)
Black Caribbean x bottom quintile	-.050 (.136)
Black Caribbean x second lowest quintile	.159 (.148)

Black Caribbean x second highest quintile	-.168 (.154)
Black Caribbean x top quintile	-.153 (.148)
Black African x bottom quintile	.017 (.108)
Black African x second lowest quintile	-.004 (.121)
Black African x second highest quintile	.090 (.115)
Black African x top quintile	.032 (.115)
Intercept	.103 (.013)***

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at reading: controlled for GOR, then limited to disadvantaged and ethnic minority strata (see Figure 12)

SEN model	**Whole sample, no control (n = 5046)**	**Whole sample, GOR control (n = 5045)**	**Ethnic / disadvantaged strata, GOR control (n = 2649)**
SEN (ref = no SEN)	-.117 (.013)***	-.118 (.015)***	-.097 (.020)***
North East(*)		.091 (.029)**	.087 (.034)**
North West(*)		.048 (.022)**	.007 (.027)
Yorks and Humber(*)		.053 (.022)**	.014 (.027)
East Midlands(*)		.093 (.023)***	.059 (.030)*
West Midlands(*)		.055 (.023)**	.025 (.028)
East England(*)		.053 (.022)**	.043 (.032)
South East (*)		.104 (.020)***	.066 (.029)**
South West(*)		.026 (.026)	-.009 (.034)
Word Reading score	.010 (.000)***	.010 (.000)***	.009 (.000)***
Intercept	-.695 (.021)***	-.644 (.030)***	-.589 (.038)***

(*)Ref = London

Income model	**Whole sample, no control (n =)**	**Whole sample, GOR control (n = 5094)**	**Ethnic / disadvantaged strata, GOR control (n = 2673)**
Low income (ref = higher income)	-.110 (.012)***	-.107 (.012)***	-.085 (.015)***
North East(*)		.098 (.029)**	.091 (.033)**
North West(*)		.053 (.021)**	.013 (.027)
Yorks and Humber(*)		.060 (.022)**	.023 (.026)
East Midlands(*)		.084 (.023)***	.050 (.030)*
West Midlands(*)		.054 (.022)**	.022 (.028)
East England(*)		.043 (.022)*	.042 (.031)
South East (*)		.087 (.020)***	.055 (.029)**
South West(*)		.016 (.023)	-.015 (.034)
Word Reading score	.010 (.000)***	.010 (.000)***	.010 (.000)***
Intercept	-.623 (.022)***	-.688 (.027)***	-.625 (.034)***

(*)Ref = London

Language model	**Whole sample, no control (n = 5098)**	**Whole sample, GOR control (n = 5097)**	**Ethnic / disadvantaged strata, GOR control (n = 2676)**
Other languages (ref = English only)	-.129 (.016)***	-.120 (.016)***	-.102 (.018)***
North East / Yorks and Humber(*)		.050 (.020)**	.024 (.024)

North West(*)		.024 (.022)	-.019 (.027)
East Midlands(*)		.063 (.024)**	.027 (.030)
West Midlands(*)		.035 (.023)	.004 (.028)
East England(*)		.025 (.023)	.026 (.032)
South East / South West (*)		.039 (.019)**	.001 (.026)
Word Reading score	0.11 (.000)***	.011 (.000)***	.010 (.000)***
Intercept	-.809 (.025)***	-.720 (.027)***	-.653 (.033)***

(*)Ref = London

Gender model	**Whole sample, no control (n = 5098)**	**Whole sample, GOR control (n = 5097)**	**Ethnic / disadvantaged strata, GOR control (n = 2676)**
Boy (ref = girl)	-.044 (.011)***	-.042 (.011)***	-.032 (.015)**
North East(*)		.094 (.029)**	.090 (.034)**
North West(*)		.053 (.022)**	.009 (.027)
Yorks and Humber(*)		.060 (.022)**	.019 (.026)
East Midlands(*)		.094 (.023)***	.055 (.030)*
West Midlands(*)		.057 (.023)**	.021 (.028)
East England(*)		.057 (.022)**	.048 (.032)
South East (*)		.102 (.020)***	.063 (.029)**
South West(*)		.030 (.023)	-.009 (.034)
Word Reading score	.010 (.000)***	.011 (.000)***	.010 (.000)***
Intercept	-.707 (.021)***	-.733 (.027)***	-.665 (.034)

(*)Ref = London

Ethnicity model	**Whole sample, no control (n = 5052)**	**Whole sample, GOR control (n = 5051)**	**Ethnic / disadvantaged strata, GOR control (n = 2643)**
Indian (ref = White)	-.083 (.032)**	-.078 (.032)**	-.037 (.036)
Pakistani (ref = White)	-.177 (.024)***	-.180 (.024)***	-.158 (.025)***
Bangladeshi (ref = White)	-.144 (.042)**	-.136 (.042)**	-.109 (.044)**
Black Caribbean (ref = White)	-.095 (.047)**	-.081 (.048)*	-.069 (.050)
Black African (ref = White)	-.187 (.037)***	-.169 (.038)***	-.160 (.040)***
North East / North West / Yorks and Humber(*)		.036 (.020)*	.004 (.024)
East Midlands / West Midlands(*)		.041 (.021)**	.004 (.026)
East England / South East / South West (*)		.023 (.019)	-.003 (.025)
Word Reading score	.011 (.000)***	.011 (.000)***	.010 (.000)***
Intercept	-.677 (.021)***	-.709 (.028)***	-.634 (.035)***

(*)Ref = London

Biases in teacher judgements of whether pupils are of 'below average' ability and attainment at reading: controlled for GOR, then limited to disadvantaged and ethnic minority strata (see Figure 13)

<u>SEN model</u>	**Whole sample, no control (n = 5046)**	**Whole sample, GOR control (n = 5045)**	**Ethnic / disadvantaged strata, GOR control (n = 2649)**
SEN (ref = no SEN)	.327 (.016)***	.327 (.011)***	.319 (.016)***
North East(*)		-.048 (.022)**	-.061 (.027)**
North West(*)		-.015 (.017)	.006 (.023)
Yorks and Humber(*)		-.014 (.017)	-.011 (.022)
East Midlands(*)		-.052 (.018)**	-.030 (.025)
West Midlands(*)		-.024 (.017)	-.003 (.023)
East England(*)		-.039 (.017)**	-.040 (.026)
South East (*)		-.050 (.016)**	-.010 (.023)
South West(*)		-.015 (.018)	.005 (.028)
Word Reading score	-.006 (.000)***	-.006 (.000)***	-.007 (.000)***
Intercept	1.174 (.016)***	.878 (.023)***	.903 (.031)***

(*)Ref = London

<u>Income model</u>	**Whole sample, no control (n =)**	**Whole sample, GOR control (n = 5094)**	**Ethnic / disadvantaged strata, GOR control (n = 2673)**
Low income (ref = higher income)	.092 (.010)***	.091 (.010)***	.086 (.013)***
North East(*)		-.071 (.024)**	-.081 (.029)**
North West(*)		-.026 (.018)	-.006 (.023)
Yorks and Humber(*)		-.040 (.018)**	-.043 (.023)*
East Midlands(*)		-.055 (.019)**	-.028 (.026)
West Midlands(*)		-.036 (.019)*	-.017 (.024)
East England(*)		-.045 (.018)**	-.060 (.027)**
South East (*)		-.033 (.017)**	.002 (.025)
South West(*)		-.008 (.019)	.022 (.029)
Word Reading score	-.008 (.000)***	-.008 (.000)***	-.009 (.000)***
Intercept	1.103 (.018)***	1.234 (.022)***	1.252 (.029)***

(*)Ref = London

<u>Language model</u>	**Whole sample, no control (n = 5098)**	**Whole sample, GOR control (n = 5097)**	**Ethnic / disadvantaged strata, GOR control (n = 2676)**
Other languages (ref = English only)	.074 (.013)***	.070 (.014)***	.076 (.016)***
North East / Yorks and Humber(*)		-.035 (.017)**	-.038 (.021)*
North West(*)		-.009 (.018)	.018 (.024)
East Midlands(*)		-.047 (.020)**	-.013 (.027)
West Midlands(*)		-.025 (.019)	-.004 (.024)
East England(*)		-.039 (.019)**	-.050 (.028)*
South East / South West(*)		-.015 (.016)	.028 (.023)
Word Reading score	-.009 (.000)***	-.009 (.000)***	-.009 (.000)***

Intercept	1.228 (.020)***	1.182 (.022)***	1.202 (.029)***

(*)Ref = London

<u>Gender model</u>	**Whole sample, no control (n = 5098)**	**Whole sample, GOR control (n = 5097)**	**Ethnic / disadvantaged strata, GOR control (n = 2676)**
Boy (ref = girl)	.050 (.009)***	.049 (.009)***	.049 (.013)***
North East(*)		-.067 (.024)**	-.078 (.029)**
North West(*)		-.025 (.018)	-.002 (.023)
Yorks and Humber(*)		-.038 (.018)**	-.035 (.023)
East Midlands(*)		-.063 (.019)*	-.033 (.026)
West Midlands(*)		-.037 (.019)**	-.016 (.024)
East England(*)		-.057 (.018)**	-.066 (.027)
South East (*)		-.044 (.017)**	-.005 (.025)
South West(*)		-.019 (.019)	.017 (.029)
Word Reading score	-.009 (.000)***	-.009 (.000)***	-.009 (.000)***
Intercept	1.178 (.017)***	1.172 (.022)***	1.194 (.030)***

(*)Ref = London

<u>Ethnicity model</u>	**Whole sample, no control (n = 5052)**	**Whole sample, GOR control (n = 5051)**	**Ethnic / disadvantaged strata, GOR control (n = 2643)**
Indian (ref = White)	.038 (.026)	.039 (.027)	.032 (.031)
Pakistani (ref = White)	.080 (.020)***	.081 (.020)***	.078 (.022)***
Bangladeshi (ref = White)	.152 (.035)***	.151 (.035)***	.155 (.038)***
Black Caribbean (ref = White)	.102 (.039)**	.099 (.040)**	.087 (.044)**
Black African (ref = White)	.159 (.030)***	.153 (.032)***	.167 (.035)***
North East / North West / Yorks and Humber (*)		-.007 (.017)	.002 (.021)
East Midlands / West Midlands (*)		-.019 (.017)	.010 (.022)
East England / South East / South West (*)		-.004 (.016)	-.023 (.022)
Word Reading score	-.009 (.000)***	-.009 (.000)***	-.009 (.000)***
Intercept	1.150 (.017)***	1.158 (.023)***	1.175 (.030)***

(*)Ref = London

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at maths: controlled for GOR, then limited to disadvantaged and ethnic minority strata (see Figure 14)

<u>SEN model</u>	**Whole sample, no control (n = 5034)**	**Whole sample, GOR control (n = 5033)**	**Ethnic / disadvantaged strata, GOR control (n =)**
SEN (ref = no SEN)	-.180 (.015)***	-.183 (.015)***	-.164 (.020)***

North East(*)		-.016 (.032)	-.034 (.036)
North West(*)		-.021 (.024)	-.040 (.029)
Yorks and Humber(*)		-.005 (.024)	-.037 (.028)
East Midlands(*)		-.016 (.025)	-.068 (.033)**
West Midlands(*)		-.024 (.025)	-.047 (.030)
East England(*)		-.007 (.024)	-.010 (.034)
South East (*)		.059 (.022)**	.005 (.031)
South West(*)		-.058 (.025)**	-.106 (.037)**
Progress in Maths score	.037 (.001)***	.037 (.001)***	.034).001)***
Intercept	-.413 (.020)***	-.228 (.027)***	-.177 (.034)***

(*)Ref = London

Income model	**Whole sample, no control (n = 5082)**	**Whole sample, GOR control (n = 5082)**	**Ethnic / disadvantaged strata, GOR control (n = 2670)**
Low income (ref = higher income)	-.095 (.014)***	-.094 (.014)***	-.066 (.017)***
North East(*)		-.013 (.032)	-.035 (.036)
North West(*)		-.018 (.024)	-.036 (.029)
Yorks and Humber(*)		.004 (.024)	-.025 (.029)
East Midlands(*)		-.030 (.026)	-.085 (.033)**
West Midlands(*)		-.027 (.025)	-.049 (.030)
East England(*)		-.017 (.024)	-.016 (.034)
South East (*)		.039 (.022)*	-.008 (.031)
South West(*)		-.068 (.025)**	-.113 (.037)**
Progress in Maths score	.040 (.001)***	.040 (.001)***	.037 (.001)***
Intercept	-.308 (.022)***	-.301 (.027)***	-.249 (.033)***

(*)Ref = London

Language model	**Whole sample, no control (n = 5086)**	**Whole sample, GOR control (n = 5085)**	**Ethnic / disadvantaged strata, GOR control (n = 2673)**
Other languages (ref = English only)	-.021 (.018)	-.022 (.018)	-.010 (.020)
North East / Yorks and Humber(*)		-.005 (.023)	-.031 (.026)
North West(*)		-.025 (.024)	-.042 (.030)
East Midlands(*)		-.029 (.026)	-.086 (.033)**
West Midlands(*)		-.030 (.025)	-.052 (.030)*
East England(*)		-.012 (.025)	-.014 (.035)
South East / South West (*)		.004 (.021)	-.046 (.028)
Progress in Maths score	.041 (.001)***	.041 (.001)***	.038 (.001)***
Intercept	-.381 (.024)***	-.350 (.027)***	-.294 (.033)***

(*)Ref = London

Gender model	**Whole sample, no control (n = 5086)**	**Whole sample, GOR control (n = 5085)**	**Ethnic / disadvantaged strata, GOR control (n = 2673)**

Boy (ref = girl)	.049 (.012)***	.050 (.012)***	.035 (.016)**
North East(*)		-.012 (.032)	-.033 (.036)
North West(*)		-.019 (.034)	-.039 (.029)
Yorks and Humber(*)		.005 (.024)	-.025 (.029)
East Midlands(*)		-.021 (.026)	-.081 (.033)**
West Midlands(*)		-.025 (.0250	-.049 (.030)
East England(*)		-.006 (.024)	-.012 (.034)
South East (*)		.053 (.022)**	.000 (.031)
South West(*)		-.056 (.025)**	-.110 (.037)**
Progress in Maths score	.042 (.001)***	.042 (.001)***	.038 (.001)***
Intercept	-.341 (.021)***	-.389 (.026)***	-.319 (.032)***

(*)Ref = London

Ethnicity model	**Whole sample, no control (n = 5040)**	**Whole sample, GOR control (n = 5039)**	**Ethnic / disadvantaged strata, GOR control (n = 2640)**
Indian (ref = White)	-.010 (.035)	-.011 (.036)	.012 (.039)
Pakistani (ref = White)	-.043 (.027)	-.041 (.027)	-.027 (.028)
Bangladeshi (ref = White)	.049 (.046)	.043 (.047)	.047 (.048)
Black Caribbean (ref = White)	-.135 (.052)**	-.145 (.053)**	-.144 (.055)**
Black African (ref = White)	-.094 (.041)**	-.107 (.043)**	-.101 (.044)**
North East / North West / Yorks and Humber(*)		-.023 (.022)	-.050 (.027)*
East Midlands / West Midlands(*)		-.040 (.023)*	-.082 (.028)**
East England / South East / South West(*)		-.014 (.021)	-.055 (.027)**
Progress in Maths score	.041 (.001)***	.041 (.001)***	.038 (.001)***
Intercept	-.357 (.021)***	-.337 (.028)***	-.270 (.035)***

(*)Ref = London

Biases in teacher judgements of whether pupils are of 'below average' ability and attainment at maths: controlled for GOR, then limited to disadvantaged and ethnic minority strata (see Figure 15)

SEN model	**Whole sample, no control (n = 5034)**	**Whole sample, GOR control (n = 5033)**	**Ethnic / disadvantaged strata, GOR control (n = 2646)**
SEN (ref = no SEN)	.345 (.012)***	.346 (.012)***	.327 (.017)***
North East(*)		-.012 (.025)	-.008 (.030)
North West(*)		.028 (.018)	.013 (.024)
Yorks and Humber(*)		.003 (.018)	.014 (.024)
East Midlands(*)		-.002 (.020)	.011 (.027)
West Midlands(*)		-.002 (.020)	.009 (.025)
East England(*)		-.002 (.019)	-.015 (.028)
South East (*)		-.028 (.017)	.013 (.026)

South West(*)		.002 (.019)	.003 (.030)
Progress in Maths score	-.024 (.000)***	-.024 (.001)***	-.025 (.001)***
Intercept	.912 (.016)***	.568 (.021)	.586 (.028)***

(*)Ref = London

Income model	**Whole sample, no control (n = 5082)**	**Whole sample, GOR control (n = 5082)**	**Ethnic / disadvantaged strata, GOR control (n = 2670)**
Low income (ref = higher income)	.073 (.011)***	.073 (.011)***	.063 (.015)***
North East(*)		-.017 (.026)	-.005 (.032)
North West(*)		.027 (.020)	.013 (.026)
Yorks and Humber(*)		-.014 (.020)	-.007 (.025)
East Midlands(*)		.014 (.021)	.038 (.029)
West Midlands(*)		-.002 (.021)	.010 (.027)
East England(*)		.002 (.020)	-.014 (.030)
South East (*)		-.003 (.018)	.035 (.027)
South West(*)		.015 (.021)	.029 (.032)
Progress in Maths score	-.032 (.000)***	-.032 (.001)***	-.033 (.001)***
Intercept	.776 (.018)***	.775 (.022)***	.787 (.030)***

(*)Ref = London

Language model	**Whole sample, no control (n = 5086)**	**Whole sample, GOR control (n = 5085)**	**Ethnic / disadvantaged strata, GOR control (n = 2673)**
Other languages (ref = English only)	-.008 (.014)	-.008 (.015)	-.012 (.018)
North East / Yorks and Humber(*)		-.015 (.019)	-.006 (.023)
North West(*)		.026 (.020)	.012 (.026)
East Midlands(*)		.006 (.022)	.031 (.029)
West Midlands(*)		-.004 (.021)	.009 (.027)
East England(*)		-.010 (.021)	-.021 (.030)
South East / South West (*)		-.009 (.017)	.023 (.025)
Progress in Maths score	-.033 (.001)***	-.034 (.001)	-.034 (.001)***
Intercept	.815 (.020)***	.826 (.022)***	.840 (.029)***

(*)Ref = London

Gender model	**Whole sample, no control (n = 5086)**	**Whole sample, GOR control (n = 5085)**	**Ethnic / disadvantaged strata, GOR control (n = 2673)**
Boy (ref = girl)	.018 (.010)*	.018 (.010)*	.018 (.014)
North East(*)		-.014 (.027)	-.003 (.032)
North West(*)		.028 (.020)	.015 (.026)
Yorks and Humber(*)		-.013 (.020)	-.002 (.025)
East Midlands(*)		.008 (.021)	.036 (.029)

West Midlands(*)		-.003 (.021)	.012 (.027)
East England(*)		-.008 (.020)	-.018 (.030)
South East (*)		-.013 (.018)	.029 (.027)
South West(*)		.006 (.021)	.026 (.032)
Progress in Maths score	-.033 (.000)***	-.034 (.001)***	-.034 (.001)***
Intercept	.830 (.017)***	.813 (.022)***	.824 (.028)***

(*)Ref = London

<u>Ethnicity model</u>	**Whole sample, no control (n = 5040)**	**Whole sample, GOR control (n = 5039)**	**Ethnic / disadvantaged strata, GOR control (n = 2640)**
Indian (ref = White)	-.016 (.028)	-.013 (.030)	-.013 (.034)
Pakistani (ref = White)	.004 (.022)	.001 (.022)	-.012 (.024)
Bangladeshi (ref = White)	-.046 (.038)	-.042 (.039)	-.049 (.042)
Black Caribbean (ref = White)	.170 (.043)***	.175 (.044)***	.168 (.048)***
Black African (ref = White)	.053 (.034)	.060 (.035)	.076 (.039)**
North East / North West / Yorks and Humber(*)		.017 (.018)	.021 (.023)
East Midlands / West Midlands(*)		.014 (.019)	.037 (.025)
East England / South East / South West(*)		.007 (.017)	.031 (.043)
Progress in Maths score	-.033 (.000)***	-.033 (.001)***	-.034 (.001)***
Intercept	.817 (.017)***	.806 (.023)***	.813 (.031)***

(*)Ref = London

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at reading: controlled for income and GOR (disadvantaged and ethnic minority strata) (see Figure 16)

<u>SEN model</u>	**Whole sample, no control (n = 5046)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2646)**
SEN (ref = no SEN)	-.117 (.013)***	-.088 (.020)***
North East(*)		.086 (.033)**
North West(*)		.010 (.027)
Yorks and Humber(*)		.016 (.026)
East Midlands(*)		.052 (.030)*
West Midlands(*)		.025 (.028)
East England(*)		.038 (.032)
South East (*)		.056 (.029)**
South West(*)		-.015 (.034)
Low income (ref = higher income)		-.082 (.016)***
Word Reading score	.010 (.000)***	.009 (.000)***
Intercept	-.695 (.021)***	-.619 (.038)***

(*)Ref = London

Language model	**Whole sample, no control (n = 5098)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2673)**
Other languages (ref = English only)	-.129 (.016)***	-.081 (.019)***
North East / Yorks and Humber(*)		.028 (.024)
North West(*)		-.011 (.027)
East Midlands(*)		.028 (.030)
West Midlands(*)		.007 (.028)
East England(*)		.025 (.032)
South East / South West(*)		.002 (.026)
Low income (ref = higher income)		-.067 (.016)***
Word Reading score	0.11 (.000)***	.010 (.000)***
Intercept	-.809 (.025)***	-.676 (.034)***

(*)Ref = London

Gender model	**Whole sample, no control (n = 5098)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2673)**
Boy (ref = girl)	-.044 (.011)***	-.033 (.015)**
North East(*)		.088 (.033)**
North West(*)		.012 (.027)
Yorks and Humber(*)		.021 (.026)
East Midlands(*)		.048 (.030)
West Midlands(*)		.020 (.028)
East England(*)		.041 (.031)
South East (*)		.052 (.029)*
South West(*)		-.015 (.034)
Low income (ref = higher income)		-.086 (.015)***
Word Reading score	.010 (.000)***	.010 (.000)***
Intercept	-.707 (.021)***	-.686 (.034)***

(*)Ref = London

Ethnicity model	**Whole sample, no control (n = 5052)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2641)**
Indian (ref = White)	-.083 (.032)**	-.032 (.035)
Pakistani (ref = White)	-.177 (.024)***	-.134 (.027)***
Bangladeshi (ref = White)	-.144 (.042)**	-.079 (.045)*
Black Caribbean (ref = White)	-.095 (.047)**	-.051 (.050)
Black African (ref = White)	-.187 (.037)***	-.140 (.041)**
North East / North West / Yorks and Humber(*)		.009 (.024)
East Midlands / West Midlands(*)		.007 (.026)

East England / South East / South West(*)		.002 (.025)
Low income (ref = higher income)		-.056 (.017)**
Word Reading score	.011 (.000)***	.010 (.000)***
Intercept	-.677 (.021)***	-.600 (.036)***

(*)Ref = London

Biases in teacher judgements of whether pupils are of 'below average' ability and attainment at reading: controlled for income and GOR (disadvantaged and ethnic minority strata) (see Figure 17)

<u>SEN model</u>	**Whole sample, no control (n = 5046)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2646)**
SEN (ref = no SEN)	.327 (.016)***	.312 (.016)***
North East(*)		-.061 (.027)**
North West(*)		.003 (.022)
Yorks and Humber(*)		-.015 (.022)
East Midlands(*)		-.025 (.025)
West Midlands(*)		-.003 (.025)
East England(*)		-.036 (.026)
South East (*)		-.003 (.023)
South West(*)		.010 (.027)
Low income (ref = higher income)		.065 (.013)***
Word Reading score	-.006 (.000)***	-.007 (.000)***
Intercept	1.174 (.016)***	.927 (.031)***

(*)Ref = London

<u>Language model</u>	**Whole sample, no control (n = 5098)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2673)**
Other languages (ref = English only)	.074 (.013)***	.051 (.017)**
North East / Yorks and Humber(*)		-.045 (.021)**
North West(*)		.008 (.024)
East Midlands(*)		-.015 (.026)
West Midlands(*)		-.008 (.024)
East England(*)		-.050 (.028)*
South East / South West(*)		.026 (.025)
Low income (ref = higher income)		.075 (.014)***
Word Reading score	-.009 (.000)***	-.009 (.000)***
Intercept	1.228 (.020)***	1.231 (.029)***

(*)Ref = London

<u>Gender model</u>	**Whole sample, no control (n = 5098)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2673)**
Boy (ref = girl)	.050 (.009)***	.050 (.013)***
North East(*)		-.077 (.029)**

North West(*)		-.006 (.023)
Yorks and Humber(*)		-.040 (.023)*
East Midlands(*)		-.025 (.026)
West Midlands(*)		-.015 (.024)
East England(*)		-.059 (.027)**
South East (*)		.006 (.025)
South West(*)		.022 (.029)
Low income (ref = higher income)		.088 (.013)***
Word Reading score	-.009 (.000)***	-.009 (.000)***
Intercept	1.178 (.017)***	1.216 (.030)***

(*)Ref = London

<u>Ethnicity model</u>	**Whole sample, no control (n = 5052)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2641)**
Indian (ref = White)	.038 (.026)	.026 (.031)
Pakistani (ref = White)	.080 (.020)***	.046 (.023)**
Bangladeshi (ref = White)	.152 (.035)***	.119 (.039)***
Black Caribbean (ref = White)	.102 (.039)**	.065 (.044)
Black African (ref = White)	.159 (.030)***	.144 (.035)***
North East / North West / Yorks and Humber(*)		-.005 (.021)
East Midlands / West Midlands(*)		.007 (.022)
East England / South East / South West(*)		.021 (.022)
Low income (ref = higher income)		.066 (.014)***
Word Reading score	-.009 (.000)***	-.009 (.000)***
Intercept	1.150 (.017)***	1.136 (.031)***

(*)Ref = London

Biases in teacher judgements of whether pupils are of 'above average' ability and attainment at maths: controlled for income and GOR (disadvantaged and ethnic minority strata) (see Figure 18)

<u>SEN model</u>	**Whole sample, no control (n = 5034)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2643)**
SEN (ref = no SEN)	-.180 (.015)***	-.159 (.020)***
North East(*)		-.034 (.036)
North West(*)		-.038 (.029)
Yorks and Humber(*)		-.036 (.029)
East Midlands(*)		-.071 (.033)**
West Midlands(*)		-.046 (.030)
East England(*)		-.014 (.034)
South East (*)		.000 (.031)
South West(*)		-.109 (.037)**
Low income (ref =		-.055 (.017)**

higher income)		
Progress in Maths score	.037 (.001)***	.033 (.002)***
Intercept	-.413 (.020)***	-.194 (.035)***

(*)Ref = London

Language model	**Whole sample, no control (n = 5086)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2670)**
Other languages (ref = English only)	-.021 (.018)	.009 (.021)
North East(*) / Yorks and Humber		-.026 (.027)
North West(*)		-.034 (.030)
East Midlands(*)		-.083 (.033)**
West Midlands(*)		-.048 (.030)
East England(*)		-.014 (.035)
South East / South West (*)		-.045 (.028)
Low income (ref = higher income)		-.069 (.018)***
Progress in Maths score	.041 (.001)***	.037 (.001)***
Intercept	-.381 (.024)***	-.321 (.034)***

(*)Ref = London

Gender model	**Whole sample, no control (n = 5086)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2670)**
Boy (ref = girl)	.049 (.012)***	.036 (.016)**
North East(*)		-.032 (.036)
North West(*)		-.036 (.029)
Yorks and Humber(*)		-.024 (.029)
East Midlands(*)		-.084 (.033)**
West Midlands(*)		-.048 (.030)
East England(*)		-.016 (.034)
South East (*)		-.006 (.031)
South West(*)		-.113 (.037)**
Low income (ref = higher income)		-.066 (.017)***
Progress in Maths score	.042 (.001)***	.037 (.001)***
Intercept	-.341 (.021)***	-.335 (.032)***

(*)Ref = London

Ethnicity model	**Whole sample, no control (n = 5040)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2638)**
Indian (ref = White)	-.010 (.035)	.017 (.039)
Pakistani (ref = White)	-.043 (.027)	.000 (.030)
Bangladeshi (ref = White)	.049 (.046)	.079 (.048)
Black Caribbean (ref = White)	-.135 (.052)**	-.121 (.055)**

Black African (ref = White)	-.094 (.041)**	-.079 (.044)*
North East / North West / Yorks and Humber(*)		-.043 (.026)
East Midlands / West Midlands(*)		-.076 (.028)**
East England / South East / South West(*)		-.053 (.028)*
Low income (ref = higher income)		-.069 (.018)***
Progress in Maths score	.041 (.001)***	.037 (.001)***
Intercept	-.357 (.021)***	-.233 (.036)**

(*)Ref = London

Biases in teacher judgements of whether pupils are of 'below average' ability and attainment at maths: controlled for income and GOR (disadvantaged and ethnic minority strata) (see Figure 19)

SEN model	**Whole sample, no control (n = 5034)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2643)**
SEN (ref = no SEN)	.345 (.012)***	.323 (.017)***
North East(*)		-.008 (.030)
North West(*)		.012 (.024)
Yorks and Humber(*)		.012 (.024)
East Midlands(*)		.013 (.027)
West Midlands(*)		.008 (.025)
East England(*)		-.012 (.028)
South East (*)		.017 (.026)
South West(*)		.005 (.030)
Low income (ref = higher income)		.042 (.014)**
Progress in Maths score	-.024 (.000)***	-.024 (.001)***
Intercept	.912 (.016)***	.599 (.029)***

(*)Ref = London

Language model	**Whole sample, no control (n = 5086)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2670)**
Other languages (ref = English only)	-.008 (.014)	-.033 (.018)*
North East / Yorks and Humber(*)		-.013 (.023)
North West(*)		.003 (.026)
East Midlands(*)		.028 (.029)
West Midlands(*)		.004 (.027)
East England(*)		-.022 (.030)
South East / South West (*)		.022 (.025)
Low income (ref = higher income)		.070 (.015)***

Progress in Maths score	-.033 (.001)***	-.033 (.001)
Intercept	.815 (.020)***	.870 (.030)***

(*)Ref = London

Gender model	**Whole sample, no control (n = 5086)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2670)**
Boy (ref = girl)	.018 (.010)*	.018 (.014)
North East(*)		-.003 (.032)
North West(*)		.013 (.026)
Yorks and Humber(*)		-.006 (.025)
East Midlands(*)		.038 (.023)
West Midlands(*)		.010 (.027)
East England(*)		-.014 (.030)
South East (*)		.036 (.027)
South West(*)		.029 (.032)
Low income (ref = higher income)		.063 (.015)***
Progress in Maths score	-.033 (.000)***	-.033 (.001)***
Intercept	.830 (.017)***	.840 (.029)***

(*)Ref = London

Ethnicity model	**Whole sample, no control (n = 5040)**	**Ethnic / disadvantaged strata, GOR and income controls (n = 2638)**
Indian (ref = White)	-.016 (.028)	-.019 (.034)
Pakistani (ref = White)	.004 (.022)	-.041 (.025)
Bangladeshi (ref = White)	-.046 (.038)	-.080 (.043)*
Black Caribbean (ref = White)	.170 (.043)***	.145 (.048)**
Black African (ref = White)	.053 (.034)	.055 (.039)
North East / North West / Yorks and Humber(*)		.014 (.023)
East Midlands / West Midlands(*)		.032 (.025)
East England / South East / South West(*)		.029 (.024)
Low income (ref = higher income)		.067 (.016)***
Progress in Maths score	-.033 (.000)***	-.033 (.001)***
Intercept	.817 (.017)***	.779 (.032)***

(*)Ref = London

Annex E: Comparisons by GOR at wave four interview between sample pupils and primary school pupils in England

Table 13: Percentage of pupils of each ethnicity in MCS sample and in DCSF statistics for population at (roughly) corresponding time period

	Indian		Pakistani		Bangladeshi		Black Caribbean		Black African		White	
	MCS	**DfE**	**MCS**	**DfE**	**MCS**	**DfE**	**MCS**	**DfE**	**MCS**	**DfE**	**MCS**	**DfE**
North East	0.83	0.57	0.83	1.41	0.00	0.90	0.00	0.03	0.00	0.45	97.11	93.24
North West	1.56	1.83	7.03	4.96	0.00	1.29	2.03	0.28	2.97	1.12	83.44	85.17
Yorkshire / Humber	1.01	1.37	20.44	8.46	2.01	0.82	0.00	0.38	0.00	1.04	73.20	82.63
East Midlands	6.29	3.92	0.63	1.82	0.00	0.58	0.63	0.62	0.84	1.28	85.53	85.76
West Midlands	4.51	4.28	8.83	8.38	3.57	1.84	0.94	1.81	0.38	1.88	76.32	74.06
East of England	1.05	1.20	3.33	2.13	3.51	1.13	1.05	0.51	0.88	1.53	85.96	86.51
London	7.82	5.34	3.44	4.02	5.01	5.37	6.10	6.30	12.36	12.70	48.04	44.76
South East	2.75	1.56	2.51	2.03	0.36	0.60	0.12	0.32	0.48	1.26	90.42	87.00
South West	0.20	0.67	0.00	0.36	0.20	0.32	0.20	0.26	0.00	0.63	98.62	93.01

(Values here are row percentages and do not sum to 100 because some ethnicities are omitted.)

Table 14: Percentage of pupils with each characteristic in MCS sample and in DCSF statistics for population at (roughly) corresponding time period

	Income				Language				Gender		SEN	
	MCS below 60% poverty level	**DfE known eligible for FSM**	**MCS rank**	**DfE rank**	**MCS: speaks lang.s other than English**	**DfE: First lang. other than English**	**MCS rank**	**DfE rank**	**MCS Male**	**DfE Males age 7**	**MCS any SEN**	**DfE any SEN**
North East	37.3	20.1	1	2	1.6	5.0	8	8	46.3	51.2	23.7	20.3
North West	34.9	18.6	3	4	10.7	11.0	4	4	51.7	50.9	22.8	18.7
Yorks / Humber	37.2	15.9	2	5	23.4	13.6	2	3	50.9	51.1	20.7	19.2
East Midlands	24.1	12.8	6	6	8.3	9.7	6	5	47.9	51.2	23.4	19.0
West Midlands	31.6	18.7	5	3	16.7	17.6	3	2	50.4	50.4	21.8	19.0
East of England	21.8	11.1	7	8	9.1	9.2	5	6	51.8	50.8	22.0	18.7
London	32.8	23.7	4	1	35.7	43.2	1	1	51.6	51.3	20.7	21.2
South East	19.6	10.1	9	9	8.2	8.8	7	7	48.4	51.1	26.7	20.7
South West	20.2	11.3	8	7	.6	4.2	9	9	50.2	51.1	20.4	19.4

National	28.2	16.0	13.7	15.2	50.1	51	22.6	19.7

Centre for Longitudinal Studies
Institute of Education
20 Bedford Way
London WC1H 0AL
Tel: 020 7612 6860
Fax: 020 7612 6880
Email cls@ioe.ac.uk
Web www.cls.ioe.ac.uk