



Stereotyping and ethnicity gaps in teacher assigned grades[☆]

Hester Burn, Laura Fumagalli, Birgitta Rabe^{*}

Institute for Social and Economic Research, University of Essex, UK

ARTICLE INFO

JEL classification:

I21
I24
J15

Keywords:

Educational assessments
Teacher grading
Ethnicity
Stereotyping
Covid-19

ABSTRACT

We explore the contribution of stereotyping to attainment gaps between pupils from different ethnic groups when grades are assigned by teachers. We exploit a change in assessment methods in England to compare grades based on teacher predictions to grades received through blindly marked examinations. When grades are assigned by teachers, ethnic minority pupils receive higher grades in maths and lower grades in English relative to White British pupils and compared to when grades are assigned through exams marked by external assessors. We use an extension of the Gelbach decomposition (Gelbach, 2016) to investigate whether the effects can be accounted for by differences in the levels of, or returns to, observed characteristics between years. Accounting for these differences partially reduces the grade gap changes in maths but roughly doubles the magnitude of the grade gap changes in English. Grade gap changes are also not driven by time trends or ceiling effects. We conclude that group-specific stereotyping is a convincing explanation of the results.

1. Introduction

Pupils are regularly assessed directly by teachers in educational settings. However, non-blind teacher assessments have been found to be susceptible to group-specific stereotyping, for example by ethnic group or gender (Botelho et al., 2015; Carlana, 2019). Stereotypes are beliefs that people have about groups with a particular characteristic and are often held unconsciously, including by teachers (Starck et al., 2020). When held, stereotypes may inform the ‘mental shortcuts’ that teachers use when they assess pupils (Burgess and Greaves, 2013), and can affect pupils in a number of important ways. They contribute to the formation of pupils’ own beliefs about their ability in school subjects and are instrumental for the subjects they choose (Burgess et al., 2022) as well as for their performance in tests (Lavy and Sand, 2018). In many contexts non-blind teacher assessments also directly influence pupils’ education trajectory at major branching points, for example in England where teacher grade predictions are used for university applications, in Germany where teacher assessments affect school tracking decisions, or in the US where they inform pupils’ Grade Point Averages. Whether and how grades assigned by teachers differ systematically according

to pupils’ characteristics is therefore an important consideration for education policy.

This paper examines the contribution of stereotyping to attainment gaps between pupils from different ethnic groups when grades are assigned by teachers. We focus on attainment gaps between White British pupils and pupils from the four largest non-White ethnic groups in England in national examinations at the end of compulsory full-time schooling, at age 16. In England, ethnic minority pupils are less likely to enrol into competitive university courses than their White peers with the same educational profile and preferences, driven in part by a mismatch between predicted and achieved grades (Del Bono et al., 2022). The contribution of stereotypes to teacher assessments may help to explain differences by ethnicity in important outcomes such as this.

We use data from the National Pupil Database, an administrative data set of children in state-funded schools in England containing students’ grades as they progress through school. We focus on two core subjects that are compulsory for all pupils at age 16: English and maths. We exploit a change in assessment methods during the 2020 Covid-19 pandemic, when examinations were cancelled, and teachers were instead asked to predict the grades the pupil would have received had

[☆] Financial support from UKRI ESRC under Covid-19 Rapid Response Project ES/V015222/1 and co-funding from the ESRC Research Centre on Micro-Social Change (ES/S012486/1) is gratefully acknowledged. We thank Angus Holford for his feedback and guidance. For useful comments and discussions, we also thank participants at the ISER student seminars and MiSoC Advisory Board Meeting 2022, and from many participants at the IWAAE and EALE conferences 2023. Burn acknowledges Ph.D. funding by The Economic and Social Research Council Research Centre on Micro-Social Change (ES/P00072X/1).

This work was produced using statistical data accessed via the ONS Secure Research Service. The use of this data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

^{*} Corresponding author.

E-mail address: brabe@essex.ac.uk (B. Rabe).

<https://doi.org/10.1016/j.labeco.2024.102577>

Received 13 November 2023; Received in revised form 24 May 2024; Accepted 25 May 2024

Available online 1 June 2024

0927-5371/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

the examinations taken place. We compare grade gaps in attainment between ethnic minority and White British pupils in 2019, when test scores were assigned through blindly marked examinations, and in 2020, when grades were assigned by teachers. This double difference approach captures changes in ethnicity grade gaps resulting from the change in assessment methods. Because stereotyped beliefs are unobservable in our data (as they often are for teachers themselves) it is not possible for us to directly relate any changes in outcomes between groups to teacher stereotyping. Instead, we follow the literature in ruling out as many other potential channels that may explain group-specific grade gap changes as possible, and interpret stereotyping as the most likely source of unexplained grade gap changes (Botelho et al., 2015; Burgess and Greaves, 2013).

We document that, when grades are assigned by teachers, ethnic minority pupils tend to do relatively better than White British pupils in maths and relatively worse than White British pupils in English compared to when the grades are from externally marked examinations. Grades were between 10 and 20% of a grade higher for ethnic minority compared to White British pupils in maths in 2020, when grades were assigned by teachers, compared to in previous years, whereas in English they were about 7% of a grade lower (with the exception of Black Caribbean pupils). Though small, these changes are not unimportant. Pupils who marginally fail to obtain a pass in English at age 16 are about nine percentage points less likely to enrol in an upper secondary qualification (Machin et al., 2020), and a one grade increase in maths is estimated to be associated with an increase of £14,579 in present value of lifetime earnings (Department for Education, 2021).

To assess the potential contribution of stereotyping to the ethnicity grade gap changes between 2020 and 2019, we consider other channels that are not related to teacher stereotyping but may drive the changes we observe. We start off by examining time trends in students' age 16 grades in maths and English. Next, we use an extension of the Gelbach decomposition (Gelbach, 2016) to investigate whether any changes in grade gaps can be accounted for by differences in the levels of, or returns to, observed characteristics between the 2019 and the 2020 cohorts. Third, as pupils received higher grades in 2020 on average, we complete separate decomposition analyses by prior attainment group to investigate whether any changes in grade gaps might be driven by ceiling effects.

Changes in observed characteristics and their returns between 2019 and 2020 explain part of the higher maths grades attained in 2020 by ethnic minority pupils relative to White British pupils. However, they do not explain the lower grades attained by ethnic minority pupils, relative to White British pupils, in English. In fact, our analysis suggests that accounting for these changes in observed characteristics would roughly double the relative drop in the English performance of ethnic minorities that followed the switch to teacher assessments. We also find that the grade gap changes are not driven by time trends or ceiling effects. There may remain characteristics which contribute to the ethnicity grade gap changes in 2020 but are unobservable in our data. Nonetheless, as long as no other explanations account for the ethnicity grade gaps changing in a positive direction for maths and negative direction for English, we argue that group- and subject-specific stereotyping is likely to drive at least part of the differences in performance between ethnic groups when grades are assigned by teachers.

As we will detail later, the timing of the announcement to cancel examinations in 2020 allows us to rule out that a change in teaching practices that might differentially favour pupils by ethnic background could be responsible for the differences we find between ethnicity grade gaps in 2020 and in years in which exams were externally marked. The announcement that examinations were to be cancelled in England occurred just two months before exams were due to begin, so that teaching practices remained largely constant across years. Teachers were also instructed to base their grade predictions only on pupils' work prior to the announcement rather than on any work completed after

school closures. Indeed, while it is possible that teachers' predictions were still informed by differences in pupils' home learning experiences during Covid-19, a substantial proportion of teachers actually stopped interacting with their pupils altogether after school closures (Eivers et al., 2020), and there is evidence that teachers primarily based their grades on the results of practice examinations taken a few months prior (Ofqual, 2021b).

The first contribution of this paper is to the literature on stereotyping which is concerned with explaining and documenting stereotypes in social situations. Our results diverge in two interesting ways from previous findings. First, we document a case where a specific group appears to be subject to both positive and negative stereotypes. While coexisting positive and negative stereotypes are themselves not unusual – for example, wealthy individuals might be viewed as both more assertive and more immoral (Tanjitpiyanond et al., 2022) – there is considerably less evidence showing divergent stereotypes within the same broad competency, such as academic achievement, outside the English context.¹ Both positive and negative stereotypes are important in our context, as both can be harmful if they affect pupils' choices or opportunities. Second, our results are *prima facie* inconsistent with economic approaches whereby stereotypes are held to be manifestations of statistical discrimination (Phelps, 1972) or representativeness heuristics (Bordalo et al., 2016; Esponda et al., 2023). In our case, most ethnic minority groups normally attain more highly than their White British peers in both maths and English, which is at odds with these explanations.

Second, this paper contributes to the literature in economics which attempts to identify the effects of using different forms of assessment, such as classroom-based assessments or examinations, in educational settings. The methodological difficulty that much of this literature seeks to overcome is that teachers often educate and assess the same pupils, making it impossible to distinguish whether effects are driven by assessment or teaching practices (Dee, 2005). Most papers address this difficulty by comparing 'blind (examination) and 'non-blind (teacher) assessments of the same pupil (Lindahl, 2007; Burgess and Greaves, 2013; Campbell, 2015). However this approach rests on the sometimes strong assumption that both assessment methods measure the same skills. For example, many studies compare examinations intended to provide a snapshot of pupils' performance in examination settings with teacher judgements which consider pupils' written, practical, and oral classwork over an entire academic year (Gibbons and Chevalier, 2008). Exploiting an exogenous change in national assessment methods as we do has, to our knowledge, not been done before, but does offer unique benefits. The teachers providing 'non-blind assessments in 2020 were asked to report the grade that they predicted their pupils would achieve had they taken blindly marked examinations. Although we cannot be certain that teachers were able to entirely disregard information they had about students that was irrelevant to the prediction of their exam grade, including information gathered both before and after school closures, this approach provides a level of comparability at least in intended outcome which is arguably greater than in some of the existing studies.

¹ The vast majority of international evidence on teacher stereotyping finds evidence of bias and stereotyping only in the same direction across subjects, including a negative bias against girls (Alan et al., 2018; Lavy and Sand, 2018; Lavy and Megalokonomou, 2019), against boys (Lavy, 2008; Lindahl, 2007), against immigrant pupils (Alesina et al., 2018; De Benedetto and De Paola, 2023), Black pupils (Botelho et al., 2015), or those perceived low-caste (Hanna and Linden, 2012). Exceptions are Terrier (2020), who finds a negative bias against boys only in maths and not in French, and Black and de New (2020), who find a negative bias against overweight pupils only in maths and not in literacy. It is worth noting that neither of the latter papers find a positive bias in the non-maths subject.

Third, this paper contributes to a literature which specifically examines systematic differences in teachers' assessments of pupils from minority or non-native ethnic groups. Measured effects tend to depend on the context being studied.² In England, ethnic minority pupils aged seven to 14 have consistently been found to receive lower teacher assessed grades than examination grades in English, and either similar or higher teacher assessed grades than examination grades in maths (Campbell, 2015; Burgess and Greaves, 2013; Gibbons and Chevalier, 2008).³ However, the evidence regarding pupils older than age fourteen is scarce.⁴ This is notable because ethnic minority pupils in England generally attain more highly relative to White British pupils as they get older (Dustmann et al., 2010), making it likely that teachers' perceptions of their skills and knowledge will also change. This paper, by focusing on teachers' predictions of ethnic minority pupils' performance across compulsory subjects in high-stakes examinations taken at age 16, therefore adds to a growing picture regarding the attainment and experiences of ethnic minority pupils as they progress through schooling in England. Predicted grades also have particular policy relevance in the setting studied; as teacher predictions of pupil examination performance are used by both further and higher education providers in England to compare applicants, our findings can also inform ongoing, national policy debates about current pupil assessment and university admissions procedures (Department for Education, 2022).

The paper proceeds as follows. Section 2 provides information on the institutional context. Section 3 outlines the data and sample used. Section 4 documents the ethnicity grade gaps and how they changed in 2020. Section 5 explores the role of time trends, while Section 6 uses decomposition analysis to explore the extent to which differences between the cohorts may have driven the patterns in ethnicity grade gap changes. Section 7 investigates whether ceiling effects matter for our findings. Section 8 provides robustness checks. Section 9 concludes.

2. Institutional context

2.1. School assessments in England

Pupils in England take examinations for the General Certificate of Secondary Education (GCSEs) at the end of full-time compulsory schooling, the summer of the year in which they turn 16. Pupils usually take examinations in eight or nine subjects, with maths and English

² Alesina et al. (2018), for example, find evidence of a negative grading bias for age 14 immigrant pupils in Italy, while Van Ewijk (2011) finds no evidence that pupil ethnicity directly affects the grades that teachers give age 11 pupils in the Netherlands, rather affecting teachers' expectations of those pupils.

³ Comparing teacher assessments with examination grades at age seven, Campbell (2015) finds that all non-White groups in her sample receive lower teacher assessed grades in reading but similar teacher assessed and examination grades in maths. Comparing teacher assessments with examination grades at age 11, Burgess and Greaves (2013) find that pupils from Pakistani, Black African, and Black Caribbean backgrounds are approximately twice as likely to receive lower teacher assessment grades than White British pupils in English, whereas for maths these likelihoods are much more similar to White British pupils for nearly all groups. Comparing teacher assessments with examination grades at age fourteen, Gibbons and Chevalier (2008) find that pupils from all ethnic minority groups have significantly lower teacher assessments and higher examination grades in English, but higher teacher assessments and lower examination grades in maths.

⁴ One exception is Murphy and Wyness (2020), who compare teacher predictions with examination grades that pupils receive at age 18, and find that Black and Asian pupils are more likely to be over-predicted than White British pupils. However due to data restrictions the authors are unable to differentiate between the subjects that the pupils study, which, at age 18, are largely determined by pupils' preferences.

(and science) as compulsory.⁵ The examinations are graded from one (low) to nine (high), with any scores below a one awarded a 'U' for 'ungraded'. Grade boundaries are set by a national regulatory body once the distribution of raw marks is known so that the proportion of pupils achieving each grade is roughly comparable between years. School-level averages are then reported in league tables to parents although, to disincentivise schools from prioritising pupils on the margin of achieving a pass grade (four), value-added measures are also given emphasis in national accountability frameworks. GCSE grades are highly determinate of the options available to pupils for post-16 education and training (Machin et al., 2020). For example most universities require applicants to have at least a grade four in English (either English Literature or English Language, whichever is higher) and maths, and any pupils who do not achieve this grade at age 16 are legally obliged to continue studying the subjects the following year.

There is little scope for grading bias in the English examination system. Grades for English Literature, English Language, and mathematics GCSE qualifications are entirely determined by pupils' performance in examinations taken at the end of a two-year course. The exam papers are externally and blindly marked by qualified teachers either from other schools or no longer in the profession, and identified by anonymous candidate numbers instead of names. Pupils' handwriting is visible to the external marker and could give away a group identity – for example if handwriting differed by gender – yet grading biases associated with handwriting have not been supported by existing evidence (Baird, 1998). It is possible for pupils to take examinations early and this could give rise to bias if it is more likely to occur for certain groups. However early entry accounts for an average of only 2.4% of entries for English Literature, English Language, and mathematics (Department for Education, 2020a), and is no more likely to occur for certain ethnic groups than for others. Schools are also able to send exam scripts to be regraded once results have been received, though this occurrence again is rare (0.05% of pupils in 2019 (Ofqual, 2020a)) and uncorrelated with pupils' demographic characteristics (Machin et al., 2020).

2.2. The change in assessment methods

Fig. 1 is a timeline detailing the change in assessment methods in 2020. In England, the school year begins in September. Schools were instructed to close because of the Covid-19 pandemic on 18th March 2020. The same announcement saw the cancellation of GCSE examinations scheduled for May and June that year. On 3rd April teachers received guidance indicating that they would be required to assign grades in place of the examinations. By mid-June schools were then asked to submit, for each pupil and for each subject in which pupils were entered, the grade that they judged the pupil would most likely have received had the examinations taken place. They were also asked to submit a rank order of each pupil in each subject. The grades and rankings were accompanied by evidence, mainly comprising

⁵ Although science is a compulsory subject, students can choose to take double or triple science. For triple science, students study separate GCSEs in biology, chemistry and physics, while for double science these subjects are combined and students receive a grade on a different scale. The two qualifications are therefore not comparable across students and, given this fact and the fact that we do not have prior attainment measures for science at age 11, we do not use science as a main outcome. However we do provide, in Appendix A, a graph of raw grade gaps in 2018, 2019, and 2020 for science, treating grades in double and triple science as comparable (we cannot include grades for 2017 as reformed science GCSEs began in 2018). The figure shows that the pattern of grade gap changes for science are similar to those we will show for maths in Section 4. As science and maths are both broadly STEM (as opposed to 'arts') subjects, this arguably expands and strengthens the case for group- and subject-specific stereotyping as a possible explanation for at least part of the grade gap changes that we observe for maths and English.

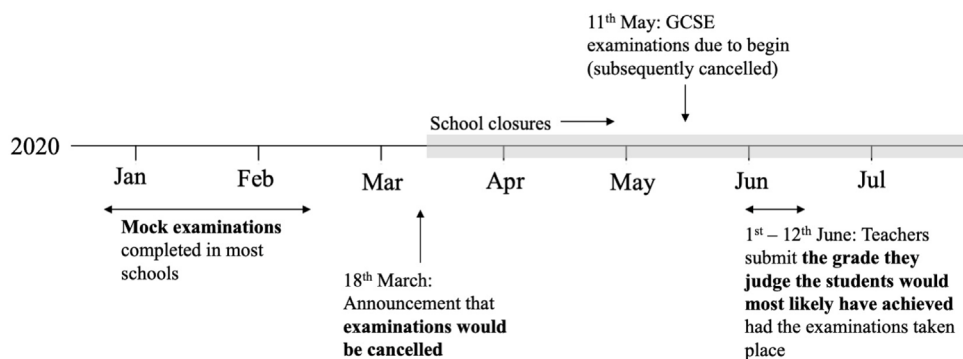


Fig. 1. Timeline showing details of the examination cancellations and teacher assigned grades in 2020.

marks and scripts from ‘mock’ (practice) examinations taken by pupils prior to school closures, often in January or February (Ofqual, 2021a). Although a statistical moderation process on these grades was initially implemented by a national regulatory body, it was later discarded due to a loss of public confidence in the process. As a result, 95% of GCSE grades received by pupils in 2020 were the predicted grades assigned directly by teachers and schools, with the remaining five percent calculated through statistical modelling (Ofqual, 2020d).

Precautions were taken to ensure that the grades that pupils received in 2020 were as comparable to those of previous cohorts as possible (Ofqual, 2020c). Schools were instructed to base their judgements on existing evidence rather than any work completed by pupils after school closures, and a survey completed by teachers shortly after the grades had been submitted showed that practice (‘mock’) examinations – which tend to be sat in exam conditions and are marked by the teacher using official marking schemes – were indeed the primary source of evidence used to inform and support their predictions (Ofqual, 2021b).⁶ Judgements made by individual teachers were also signed off by at least two members of staff – one of which was the lead teacher of the subject within the school – and head teachers (principals) were required to submit a declaration that the grades had been generated according to the guidance. Teachers were also instructed to not discuss the grades with pupils or their families and in many cases did not continue to interact with these pupils after schools were closed and examinations cancelled, instead prioritising online learning provision for other groups (Eivers et al., 2020). We therefore assume that the 2020 grades are not influenced by differences in school closure experiences between groups, or by manipulation to appease families.

3. Data and sample

We use the National Pupil Database (NPD) which contains administrative data on the universe of pupils in England who attend state-funded schools. It includes demographic information about pupils, measures of their attainment at age seven, 11, 16, and 18, and information about schools and local authorities. As all state-funded schools and examination boards in England are required to return these data by

⁶ Marks from mock examinations are kept ‘on file’ to use to facilitate progression to further education, training, or employment if the pupil is unable to sit the actual exam, so must be marked strictly. Teachers have little incentive to be inaccurate; though a lower mark may ‘encourage’ a pupil to revise more for the real exam, a lower mark would be detrimental if for any reason the mock needs to be used in place of the actual exam. In this way the grade from mock examinations should be very close to those from a real exam. It would also be reasonable to assume that any pupils who experience test anxiety are likely to be affected similarly in the mocks as in the real exams, further increasing the level of comparability between the information that teachers were using to make grade predictions and pupils’ likely performance.

law, the NPD is both accurate and highly complete. We use the cohorts of pupils in the NPD who took exams to mark the end of compulsory full-time schooling in the academic years ending in 2017, 2018, 2019 and 2020. As the change to teacher predicted grades occurred for the end of year examinations in 2020 only, we remove any grades from examinations to which pupils were entered a year early. This inevitably includes a small proportion for whom this was the only recorded grade in a given subject (about 2.5% of pupils per cohort). The resultant sample comprises 2,252,123 pupils, or about 560,000 observations for each of the four cohorts. We drop the exam results of pupils in fee-paying schools and non-mainstream schools which together comprise 6.78% of the data.⁷ We then remove pupils with missing data for any of the pupil or school characteristics we consider (12.7%) or with no recorded grade for GCSE maths or either GCSE English Literature or GCSE English Language (0.800%).⁸ This leaves a final sample size of 1,823,542 pupils in 3,380 schools across all four cohorts.

Our outcomes of interest are GCSE (age 16) grades in English and maths. There is one maths GCSE qualification in England. However, there are two GCSE English qualifications: English Language and English Literature. The vast majority of pupils in England take both. If their grades differ, the higher of the two is used in headline performance measures for schools and, by pupils, to meet performance benchmarks (Department for Education, 2020b). We follow these national conventions and use the higher of the two grades in our main analyses. Results for English Language and English Literature separately, as well as for the mean of the two, are presented as a robustness check.⁹

⁷ Fee-paying schools are not required to submit Pupil Census data so we cannot observe pupil characteristics. The non-mainstream state-funded schools that we exclude are those solely for pupils with special educational needs, in hospital, who have been removed from mainstream education due to behavioural concerns, or who have been given custodial sentences. Low proportions of these students are entered for GCSE examinations.

⁸ The vast majority of missing pupil or school characteristics are prior attainment scores at age 11. Missing grades are rare, and likely to occur because pupils are entered for different types of qualification (for example for those unlikely to reach GCSE standard), or pupils being entered for the GCSE but not showing up on the day.

⁹ In June 2020, 567,277 and 531,626 16-year-old pupils received English Language and English Literature grades respectively, which is similar in proportion to 2019 for which the figures are 546,607 and 514,191 (Joint Council for Qualifications, 2020). The overall increase in entries is due to the 2020 cohort being approximately 3% larger than that of 2019 (Ofqual, 2020b). 5.4% of pupils were entered for English Literature a year early in 2019, which explains much of the discrepancy between the subjects (Department for Education, 2020a). As our sample is restricted to the end of year examinations in the year in which pupils turn 16, such ‘early entries’ are dropped. However these differences are not likely to cause any differences in results between English Literature and English Language. First, we are primarily interested in changes between 2019 and 2020, and patterns of entries appear consistent

The ethnic groups that we consider are Pakistani and Bangladeshi, Indian, Black African, and Black Caribbean, which are the largest ethnic minority groups in England and Wales, and we compare them to White British students (Office for National Statistics, 2022).¹⁰ The additional pupil-level characteristics we consider in the decomposition analyses are gender, whether the pupil has been identified as having a special educational need (SEN), whether the pupil speaks English as an additional language (EAL), and pupils' attainment at age 11 (KS2 subject-specific point scores), which we standardise by year and subject.¹¹ As proxies for pupils' socio-economic status, we use eligibility for free school meals at any point in the last six years (FSM6) and a rank based on the proportion of all children aged 0 to 15 living in income deprived households in the pupils' local area of residence (IDACI score). At the school-level, we consider a number of characteristics including the type of school, whether the school has selective admissions, and the school region. School quality is proxied by a measure of school value-added from the end of primary school to examinations at age 16 in eight prescribed subjects (called 'Progress 8').

We present descriptive statistics for our sample, separately for each year between 2017 and 2020, in Table B.1, and separately for each ethnic group across years in Table B.2. Table B.2 shows that the pupils in our sample who are members of our ethnic minority groups of focus live, on average, in more deprived neighbourhoods than White British pupils, and are more likely to speak English as an additional language. Compared to the other ethnic minority groups, Indian pupils are also more likely to attend schools with selective admissions, and Black Caribbean pupils both less likely to speak English as an additional language and more likely to be diagnosed with a special educational need.

4. Ethnicity grade gaps

We begin by documenting raw ethnicity grade gaps at age 16 in England and how they change over time. Fig. 2 shows raw grade gaps for the four ethnic minority groups in 2017, 2018, 2019 and 2020. For each year, the bar shows the difference in the average grade that pupils received in relation to the reference group, White British pupils, in that year. Pakistani and Bangladeshi, Indian, and Black African pupils generally attain more highly than White British pupils in both English and maths, while Black Caribbean pupils attain lower grades in both subjects. Of particular interest are the differences in the 2020 gaps compared to the same differences in the preceding years. In maths, all ethnic minority groups received higher grades relative to White British pupils than they did in the preceding years, whereas in English all apart from Black Caribbean pupils received lower relative grades than in the preceding years. These grade gaps by year and subject are also reported in Panel A of Table 1. We will show in Section 6 that once we control for changes across cohorts in the levels of, and returns to, observable characteristics, grade gap changes in English for the Black Caribbean group become more aligned with those of other ethnic minority groups.

The first row of Panel B in Table 1 reports the difference between the grade gap of each ethnic minority group and White British pupils in 2020 and in 2019 (i.e. the difference between the fourth and third row of Panel A of Table 1), alongside standard errors on the difference. We call these double differences the 'ethnicity grade gap changes'. For

between these years (Joint Council for Qualifications, 2020). Second, early entrants in English Literature are proportionate to the full sample with regards to ethnicity.

¹⁰ In our analyses we also include indicators for 'Multiethnic' and 'Any Other Ethnic Group' but do not report or interpret results for these groups as each combines a very diverse group of pupils.

¹¹ For KS2 subject-specific point scores, we use the total marks in the KS2 maths tests for maths, and the marks in the KS2 English reading test for English. All of these tests are blindly and externally marked.

every subject and ethnic minority group, these changes measure how the grade gap between the ethnic minority group and White British pupils in 2020 (when grades are assigned by teachers) compares to that same gap in 2019, when grades are assigned through examinations marked by external examiners. A positive change indicates that, compared to White British pupils, an ethnic minority group receives relatively higher grades when grades as assigned by teachers than through examinations marked by external examiners; a negative change indicates that, compared to White British pupils, an ethnic minority group is given relatively lower grades when grades are assigned by teachers than through examinations marked by external examiners.

For maths, the raw ethnicity grade gap changes are positive for all ethnic minority groups. For Pakistani and Bangladeshi, Indian, and Black African pupils, who all achieve more highly relative to White British pupils in 2019, the relative grade gap changes for maths between 2019 and 2020 are 10.1%, 9.1%, and 11.5% of a grade respectively. For Black Caribbean pupils, who achieve lower than White British pupils in prior years, the relative grade gap change in 2020 is 20.1%. In other words, ethnic minority pupils received relatively higher maths grades than White British pupils in 2020 when the grades were assigned by teachers. The raw ethnicity grade gap changes for English are less consistent. For Pakistani and Bangladeshi and Black African pupils they are negative, at -6% and -10.3% of a grade respectively, and significant at the one percent level. Pakistani and Bangladeshi and Black African pupils therefore received relatively lower grades than White British pupils in 2020 compared to 2019 in English. The point estimate for Indian pupils is negative (-4.1%) though not statistically significant. For Black Caribbean pupils it is positive (8.2%).

The pattern of higher relative grades in maths and lower relative grades in English is consistent with the findings of the extant research into ethnic differences in teacher assessment and examination grades in England for younger pupils; ethnic minority pupils tend to receive lower teacher assessed grades than examination grades in English and either similar or higher teacher assessed grades than examination grades than in maths (Burgess and Greaves, 2013; Campbell, 2015; Gibbons and Chevalier, 2008). The size of the estimated grade change is not negligible. Using the standard deviations for 2020 (see Table B.5), a 10% grade change in either subject is equivalent to an effect size of approximately 0.05, which is the same as the average effect size of school-based interventions aimed at improving academic achievement in developed countries (Fryer, 2017).

5. The role of time trends

We want to ensure that the gap changes presented in Panel B of Table 1 are driven by the switch to teacher assigned grades, rather than the results of pre-existing trends. Therefore, we explore the extent to which the ethnicity grade gap changes in 2020 differed from trends in the grades received by ethnic minority pupils relative to White British pupils over time. We first regress each subject grade on an indicator for ethnicity, year, and their interaction, excluding the year 2020. This estimates separate trends in grades by subject and ethnic group for the years in which grades were assigned only through blindly-marked examinations.¹² We then use the coefficients estimated from this model to predict grades (and thereby grade gaps) in 2020. Finally, we test

¹² We make Y_i , that is a pupil's grade, depend on a set of observed characteristics as follows:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 year_i + \alpha_3 [year_i \times X_{1i}] + \eta_i$$

where X_{1i} contains dummy variables indicating whether the pupil is in any of the ethnic minority groups and $year$ is a continuous variable. We then use this model to estimate the predicted grades for all ethnic groups in the year 2020. As the grades in 2020 are based on a naive prediction, there is no variance in these predictions by group and year and we do not report a standard error.

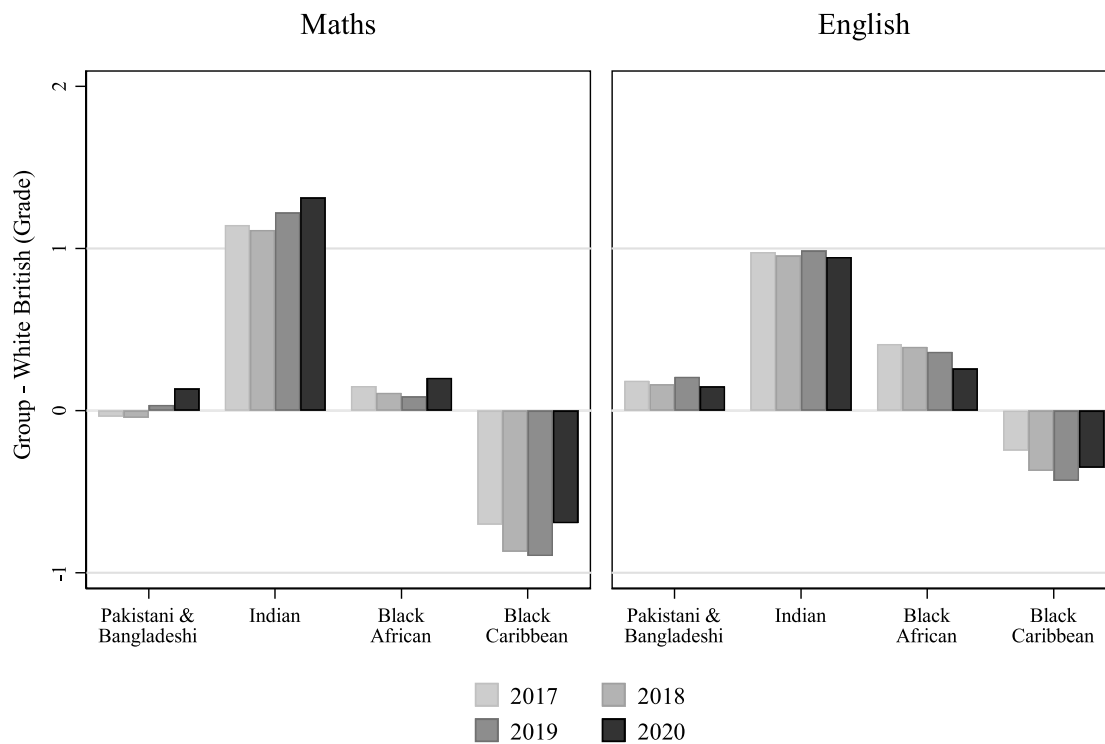


Fig. 2. Grade gaps by ethnic group, year and subject.

Notes: National Pupil Database. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both.

Table 1
Grade gaps by ethnic group, year and subject.

	Maths				English			
	P&B	I	BA	BC	P&B	I	BA	BC
Panel A: Raw ethnicity grade gaps								
Group - White British:								
2017	-0.037 (0.035)	1.145*** (0.054)	0.153*** (0.032)	-0.703*** (0.038)	0.185*** (0.033)	0.978*** (0.041)	0.412*** (0.032)	-0.247*** (0.037)
2018	-0.044 (0.036)	1.114*** (0.052)	0.110*** (0.032)	-0.870*** (0.039)	0.164*** (0.034)	0.956*** (0.044)	0.393*** (0.032)	-0.371*** (0.037)
2019	0.036 (0.037)	1.222*** (0.052)	0.089*** (0.032)	-0.895*** (0.035)	0.210*** (0.035)	0.988*** (0.044)	0.363*** (0.031)	-0.432*** (0.036)
2020	0.137*** (0.034)	1.314*** (0.048)	0.204*** (0.028)	-0.694*** (0.031)	0.149*** (0.032)	0.947*** (0.039)	0.260*** (0.027)	-0.350*** (0.030)
N	1,823,542				1,823,542			
Panel B: Ethnicity grade gap changes								
Double difference:								
2020-2019 change	0.101*** (0.023)	0.091*** (0.028)	0.115*** (0.024)	0.201*** (0.035)	-0.060*** (0.020)	-0.041 (0.028)	-0.103*** (0.023)	0.082** (0.036)
N	934,590				934,590			
Panel C: Predicted ethnicity grade gaps								
Predicted 2020	0.059	1.239	0.053	-1.015	0.212	0.984	0.340	-0.534
Actual - predicted 2020	0.078** (0.034)	0.075 (0.048)	0.151*** (0.028)	0.320*** (0.031)	-0.062** (0.032)	-0.037 (0.039)	-0.080*** (0.027)	0.184*** (0.030)
N	1,823,542				1,823,542			

Notes: National Pupil Database. 'P&B' stands for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. The reference group is White British. Standard errors in parentheses, clustered at the school-level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

whether or not these predicted 2020 grade gaps are different from those observed in 2020 (fourth row of Panel A, Table 1).

We present the results in Panel C of Table 1, separately for maths and English. The first row of the panel shows the 2020 grade gap predicted based on a time trend, and the second shows the differences between the actual 2020 grade gap and the predicted grade gap. These

differences are the raw ethnicity grade gaps in 2020 net of a time trend. Comparing the differences to the observed ethnicity grade gap changes in the first row of Panel B of Table 1 suggests no clear pattern. Accounting for a time trend in some cases exacerbates the gaps for some ethnic minority pupils, compared to White British pupils, and in other cases the gap remains unchanged or is attenuated. In the latter cases

(Pakistani and Bangladeshi pupils in maths and Black African pupils in English) the trends predict only about 20% of the actual grade gap changes that we observe, indicating that time trends do not explain the ethnicity grade gap changes we observe.

However, grade gaps in 2020 could differ because returns to characteristics other than ethnicity are different in 2020 compared to earlier years, due to the change to teacher assessment. For example, a pupil characteristic such as eligibility for free school meals may have greater returns for attainment in one subject in 2020 compared to 2019. If this characteristic is more greatly represented in a specific ethnic group, then this group will see a greater relative grade increase in 2020 due to factors other than ethnicity. Understanding the role of these factors is the focus of the next section.

6. The role of differences between cohorts

In this section we assess whether changes in pupil characteristics and their returns across years explain the grade gap changes we documented, and, if so, which characteristics mostly contribute to the results. To do so, we decompose the changes in performance gaps between 2019 and 2020 into an ‘explained part’, namely the part explained by changes in returns to observed characteristics or changes in their prevalence by group, and an ‘unexplained’ part. We use an extension of the standard Gelbach decomposition (Gelbach, 2016) that allows returns to characteristics to differ by year and, unlike the Oaxaca–Blinder decomposition (Oaxaca, 1973; Blinder, 1973), yields results that are order-invariant, that is, they do not depend on the order in which observed characteristics are added to the model. We allow both the levels of, and the returns to, observed characteristics other than ethnic group to differ across the years. We do not allow the returns to observed characteristics to differ by ethnic group as we want the ‘explained’ gap change to capture the part of the grade gap changes which can be accounted for by factors other than ethnicity and ethnicity-based stereotyping. The effects of ethnicity and stereotyping, including changes in the returns to observed characteristics that differ by ethnic group as well as other teacher stereotyping effects acting through variables that are not observed in our model, are then contained in the ‘unexplained’ part.

6.1. Decomposition methodology

We start by formalising the raw ethnicity grade gap changes as:

$$Y_i = \beta_0^{raw} + \beta_1^{raw} \mathbf{X}_{1i} + \beta_2^{raw} 2020_i + \beta_3^{raw} [2020_i \times \mathbf{X}_{1i}] + \mu_i \quad (1)$$

where Y_i indicates pupil i 's grades, \mathbf{X}_{1i} is a vector of indicators of whether pupil i is in any of the ethnic minority groups, 2020_i is an indicator for whether i is in the 2020 cohort, $[2020_i \times \mathbf{X}_{1i}]$ contains the interactions between the 2020_i indicator and the variables in \mathbf{X}_{1i} , and μ_i is an error. Here β_0^{raw} is the mean grade achieved by White British pupils in 2019, β_1^{raw} is the vector of gaps in the mean grade of each ethnic minority group compared to White British pupils in 2019, β_2^{raw} is the partial effect of the year 2020 on the mean grade for White British pupils, and β_3^{raw} is a vector of the raw ethnicity grade gap changes in 2020 compared to 2019 (the same gap changes we show in Panel B of Table 1).

To derive the unexplained ethnicity grade gap changes we can estimate the following equation:

$$Y_i = \beta_0 + \beta_1 \mathbf{X}_{1i} + \beta_2 2020_i + \beta_3 [2020_i \times \mathbf{X}_{1i}] + \beta_4 \mathbf{X}_{2i} + \beta_5 [2020_i \times \mathbf{X}_{2i}] + \varepsilon_i \quad (2)$$

where Y_i , \mathbf{X}_{1i} , 2020_i are as defined above, and \mathbf{X}_{2i} contains observable individual and school characteristics of pupil i including gender, family and neighbourhood deprivation (SES), whether the pupil has a special educational need, whether the pupil speaks English as their first language, the pupils' subject-specific prior attainment at age 11, a number of school characteristics, and a measure of school quality, $[2020_i \times \mathbf{X}_{2i}]$ is the interaction between \mathbf{X}_{2i} and the 2020_i indicator,

and ε_i is an error term. β_3 is the vector of the unexplained ethnicity grade gap changes, that is what is left of the raw ethnicity grade gap changes after controlling for the observable characteristics in \mathbf{X}_{2i} and their interactions with 2020_i , β_4 is the vector of returns to \mathbf{X}_{2i} on y_i in 2019, and β_5 is the vector of the changes in the returns to \mathbf{X}_{2i} on y_i in 2020 compared to 2019.

The difference between the raw and the conditional (unexplained) ethnicity grade gap changes ($\beta_3^{raw} - \beta_3$) are the explained grade gap changes, which are contained in the vector δ . The explained grade gap changes combine the effects of changes in the distribution of observed characteristics across ethnic groups in 2020 compared to 2019, and the effects of changes in the returns to these characteristics in 2020 compared to 2019. We can examine these combined effects and how they differ for different pupil characteristics contained in \mathbf{X}_2 by implementing an extension of the standard decomposition proposed by Gelbach (2016) in which we allow the returns to the observed characteristics to differ by year.¹³ To do this, for each characteristic k in \mathbf{X}_2 , we estimate the differences in the levels (means) of that characteristic for each ethnic group compared to White British pupils in 2019 ($\hat{I}_{2019,eth}^k$) and the changes in those differences in levels in 2020 ($\hat{I}_{2020,eth}^k$) using a set of auxiliary regressions.¹⁴ We then use $\hat{I}_{2019,eth}^k$ and $\hat{I}_{2020,eth}^k$ as well as coefficients from Eq. (2) to derive the vector of explained gap changes, $\hat{\delta}$, as follows:¹⁵

$$\hat{\delta} = \hat{\beta}_3^{raw} - \hat{\beta}_3 = \sum_{k=1}^K [\hat{I}_{2020,eth}^k \hat{\beta}_4^k + \hat{I}_{2019,eth}^k \hat{\beta}_5^k + \hat{I}_{2020,eth}^k \hat{\beta}_5^k] \quad (3)$$

where $\hat{\delta}$ is the explained part of $\hat{\beta}_3^{raw}$ made up of the sum of $\hat{I}_{2020,eth}^k \hat{\beta}_4^k$ (the changes in the differences in the levels of k in 2020 multiplied by the returns to k in 2019), $\hat{I}_{2019,eth}^k \hat{\beta}_5^k$ (the differences in the levels of k in 2019 multiplied by the changes in the returns to k in 2020), and $\hat{I}_{2020,eth}^k \hat{\beta}_5^k$ (the changes in the differences in the levels of k in 2020 multiplied by the changes in the returns to k in 2020).

$\hat{\delta}$ can then be split into components by ethnic group and pupil characteristic. Consider for example the contributions of the gender (‘male’) and socio-economic status (‘SES’) characteristics to the change in the grade gap between Pakistani and Bangladeshi (‘PB’) and White British pupils in 2020 compared to 2019. They can be written as follows:

$$\text{Male component: } \hat{\delta}_{PB}^{male} = \hat{I}_{2020,PB}^{male} \hat{\beta}_4^{male} + \hat{I}_{2019,PB}^{male} \hat{\beta}_5^{male} + \hat{I}_{2020,PB}^{male} \hat{\beta}_5^{male}$$

$$\text{SES component: } \hat{\delta}_{PB}^{SES} = \hat{I}_{2020,PB}^{SES} \hat{\beta}_4^{SES} + \hat{I}_{2019,PB}^{SES} \hat{\beta}_5^{SES} + \hat{I}_{2020,PB}^{SES} \hat{\beta}_5^{SES}$$

Here the male (SES) component shows how changes between 2019 and 2020 in the proportion of Pakistani and Bangladeshi compared to White British males (low SES pupils) and in the returns to being male (a low SES student) contribute to the observed changes in the ethnicity grade gaps in 2020 compared to 2019 in the absence of ethnicity-based stereotyping, that is when changes in the returns to the characteristics do not differ by ethnic group.

¹³ See in particular (Gelbach, 2016) Sections IV and V.B.

¹⁴ $\hat{I}_{2019,eth}^k$ and $\hat{I}_{2020,eth}^k$ can be estimated using a set of auxiliary models with each of the k characteristics in \mathbf{X}_2 acting as the dependent variable such that

$$\mathbf{X}_{2i}^k = \Gamma_{2019}^k + \Gamma_{2019,eth}^k \mathbf{X}_{1i} + \Gamma_{2020}^k 2020_i + \Gamma_{2020,eth}^k [2020_i \times \mathbf{X}_{1i}] + w_i^k$$

where k represents one of the characteristics in \mathbf{X}_2 , and w_i^k is a residual.

¹⁵ We implement the decomposition using the ‘b1x2’ Stata command based on Gelbach (2016).

Table 2
Gelbach decomposition of the change in grade gaps by ethnic group.

	Maths				English			
	P&B	I	BA	BC	P&B	I	BA	BC
Raw gap change	0.101*** (0.023)	0.091*** (0.028)	0.115*** (0.024)	0.201*** (0.035)	-0.060*** (0.020)	-0.041 (0.028)	-0.103*** (0.023)	0.082** (0.036)
Explained gap change	0.065*** (0.022)	-0.009 (0.024)	0.018 (0.021)	0.056* (0.030)	0.080*** (0.019)	0.036* (0.022)	0.075*** (0.020)	0.104*** (0.025)
<i>Amount explained by:</i>								
Male	0.001 (0.002)	0.000 (0.002)	0.002 (0.002)	0.003* (0.002)	0.003 (0.003)	0.004 (0.004)	-0.004 (0.004)	0.005 (0.005)
Socio-economic status	0.006 (0.004)	0.012*** (0.003)	0.015*** (0.005)	0.006 (0.006)	0.015*** (0.004)	0.011*** (0.003)	0.031*** (0.006)	0.020*** (0.006)
Special educational needs	0.004*** (0.001)	0.007*** (0.001)	0.004*** (0.001)	-0.003 (0.002)	0.005*** (0.002)	0.007*** (0.002)	0.006*** (0.002)	0.000 (0.004)
First language not English	0.011 (0.009)	0.008 (0.008)	0.005 (0.007)	0.000 (0.001)	0.011 (0.010)	0.009 (0.008)	0.006 (0.007)	0.000 (0.001)
Prior attainment (age 11)	0.024 (0.016)	-0.026 (0.019)	-0.015 (0.016)	0.027 (0.026)	0.002 (0.010)	0.036*** (0.014)	0.030*** (0.012)	0.047*** (0.016)
School characteristics	0.004 (0.008)	-0.008 (0.007)	0.014 (0.010)	0.018* (0.011)	0.031*** (0.009)	-0.023*** (0.008)	0.018 (0.011)	0.029** (0.012)
School value-added (lagged)	0.015* (0.009)	-0.003 (0.009)	-0.008 (0.008)	0.004 (0.010)	0.012 (0.009)	-0.008 (0.010)	-0.011 (0.008)	0.003 (0.010)
Unexplained gap change	0.036* (0.019)	0.100*** (0.020)	0.096*** (0.019)	0.145*** (0.026)	-0.140*** (0.020)	-0.077*** (0.025)	-0.179*** (0.021)	-0.022 (0.030)
N	934,590				934,590			

Notes: National Pupil Database. 'P&B' stands for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. The gap is the gap in average grades between the ethnic minority group and White British pupils. Socio-economic status includes free school meals indicator and rank based on the proportion of all children aged 0 to 15 living in income deprived households in the pupil's local area of residence (IDACI score). Prior attainment (age 11) includes subject-specific age 11 attainment score, standardised by year. School characteristics are indicators for selectivity, urban, region (9 categories), and school governance type (4 categories), and continuous measures of cohort size, proportion of pupils eligible for free school meals, and average neighbourhood deprivation (IDACI) score of the school. School value-added refers to how many average grades higher or lower the pupils in that school achieve across eight qualifying GCSE subjects compared to pupils across the country who score comparatively at age 11. Standard errors in parentheses, clustered at the school level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

6.2. Decomposition results

Table 2 reports the results for the decomposition of the changes in the grade gaps between each ethnic minority group and White British pupils in 2020 compared to 2019. The first row shows the raw ethnicity grade gap changes and is identical to that reported in the second panel of Table 1. The second row shows estimates for δ , that is the part of the ethnicity grade gap changes which are explained by differences in the levels of, and returns to, observed characteristics between the years. In other words, it shows the ethnicity grade gap changes that we would have expected in 2020, given observed changes in the characteristics and their returns between years (and assuming no changes in the differences in returns to these characteristics by ethnic group). Below this, Table 2 shows the separate contributions of different characteristics or groups of characteristics to this overall explained component.¹⁶ Finally, the bottom row shows contributions to the ethnicity grade gap changes which remain unexplained. These include changes in the returns to observed characteristics that differ by ethnic group as well as any factors unobserved in our model, including teacher stereotyping effects acting through variables that are not observed. The estimated returns can be found in Table B.4.

The decomposition results for maths are reported in the four columns on the left of Table 2. Changes in the levels of and returns to pupils' socio-economic status and diagnoses of special educational needs across groups and years contribute to explaining the ethnicity grade gap changes for most ethnic minority groups, but to a relatively small extent.¹⁷ Note that the estimated contributions of subject-specific

¹⁶ Each of the individual characteristic contributions separately can be found in Table B.3.

¹⁷ Changes in school characteristics and being male contribute positively to the ethnicity grade gap changes for Black Caribbean pupils, and in school value-added for Pakistani and Bangladeshi pupils.

previous attainment at age 11 do not reach statistical significance (see seventh row of Table 2). In fact, in the case of Indian and Black African pupils, the point estimates of these contributions are negative, in contrast with the positive raw gap changes these contributions should help explain.

Taken together, for Pakistani and Bangladeshi pupils, changes in the observed characteristics and their returns explain 6.5% of a grade (see second row of Table 2), that is around two thirds of the 10% raw maths grade gap change, with only 3.6% of a grade left unexplained (see bottom panel of Table 2). For Black Caribbean pupils, the observed characteristics and their returns explain 5.6% of a grade, that is less than a third of the over 20% of a grade raw maths grade gap change, leaving 14.5% of a grade unexplained. For Indian and Black African pupils, changes in the observed characteristics and their returns explain very little of the raw maths grade gap change, as shown by the small and statistically insignificant estimates of the explained gap changes (second row of Table 2). As a consequence, a considerable proportion of the raw grade gap changes are left unexplained, as we can see by comparing the top and the bottom panel of the second and third columns of Table 2.

The decomposition results for English are reported in the four columns on the right of Table 2. Pupils' socio-economic status, diagnoses of special educational needs, and school characteristics all contribute to explaining the ethnicity grade gaps. However, for most ethnic groups, pupil prior attainment (row 7) appears to be the most important explanatory factor. This is particularly relevant in the case of Black Caribbean pupils, where changes in prior attainment and their returns explain more than half of the observed grade gap change.¹⁸ The

¹⁸ The main characteristic that is different for the 2020 Black Caribbean cohort is a large increase in English prior attainment, both compared to the other cohorts and to other ethnic minority groups. We can see from Table B.4 that in English prior attainment is more highly rewarded in 2020 than in 2019, and this has a large effect on Black Caribbean students in 2020.

second row of [Table 2](#) shows the explained components of the grade gap, summarising the contribution of all the observed characteristics. These components are positive, larger than those estimated for maths and statistically significant for all ethnic groups. These results indicate that, given the changes in the groups' characteristics and the returns to those characteristics in 2020 compared to 2019, especially in prior attainment, and in absence of differential changes in returns by ethnic group, we would have expected all ethnic minority pupils to have achieved more highly in 2020 than in 2019 relative to White British pupils. We would have expected Pakistani and Bangladeshi pupils to have achieved relatively more highly by 8% of a grade, Indian pupils by 3.6% of a grade, Black African pupils by 7.5% of a grade, and Black Caribbean pupils by 10.4% of a grade. It is therefore noteworthy that instead we see negative raw grade gap changes for most groups, showing that the opposite was in fact the case — they achieved relatively lower.

The bottom panel of [Table 2](#) reports the unexplained gap changes, which are the differences between the ethnicity grade gap changes in English predicted by changes in the observed characteristics and their returns and the raw ethnicity grade gap changes. These unexplained components of the English grade gap changes are negative for all ethnic groups. For Pakistani and Bangladeshi, Indian and Black African pupils, the unexplained components of the English grade gap changes are highly statistically significant, and around double in magnitude than the raw grade gap changes (compare the top and the bottom panel of [Table 2](#)). This suggests that in 2020, when grades were assigned by teachers, these ethnic minority groups received lower grades in English than White British pupils when compared to the previous year despite the 2020 cohort having characteristics (previous attainment in particular) that, in absence of differential changes in returns by ethnic group or unobserved factors, would have predicted an improvement in their relative performance. Unlike for these groups, the raw 2020 grade gap change in English for Black Caribbean pupils was positive (see the last column of the top row of [Table 2](#)). However our decomposition shows that this is entirely driven by changes in observed characteristics and their returns. Once we control for these, the point estimate of the unexplained gap change in English for the Black Caribbean group becomes negative – although not statistically significant – and thus aligns more with the other groups (see bottom panel of [Table 3](#)).

In summary, [Table 2](#) shows that changes in observed characteristics and their returns explain part of why ethnic minority pupils received relatively higher grades in maths in 2020, though a substantial proportion of the raw grade gap changes remain unexplained. Changes in observed characteristics and their returns fully explain why Black Caribbean pupils received relatively higher grades in English. However, they do not explain why Pakistani and Bangladeshi, Indian and Black African pupils received relatively lower grades in English. In fact, our results suggest that had those changes not occurred, the relative drop in the English performance that followed the switch to teachers' assessment would have been roughly double in magnitude for these ethnic minority groups. Note that the direction of the effects of the explained contributions are roughly consistent across both maths and English, so we would not expect any factors that we are not able to observe to be causing the contrasting direction of the unexplained grade gap changes. We conclude that there is evidence suggesting that unexplained factors, differing by subject and likely including teacher stereotyping, had a role in determining the 2020 GCSE results.

7. The role of ceiling effects

Another way that the shift to teacher assessments might have indirectly affected the relative performance gaps across ethnic groups is through ceiling effects. Ceiling effects, a type of scale attenuation effect, are observed when there is an upper limit to the dependent variable. In our case this is the GCSE upper limit of a grade 9. Ceiling effects may in principle be relevant in our case as, as shown in [Table 3](#), the grade

Table 3
Mean grades by subject and year.

	Maths	English
2017	4.71 (2.03)	5.23 (1.90)
2018	4.75 (2.03)	5.22 (1.90)
2019	4.77 (2.03)	5.22 (1.90)
2020	5.06 (1.99)	5.40 (1.82)
N	1,823,542	1,823,542

Notes: National Pupil Database. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. Standard deviations in parentheses.

averages for both maths and English in 2020 were considerably higher than in the preceding years. Higher grades in 2020 therefore resulted in a smaller spread of grades overall (as evidenced by slightly smaller standard deviations), likely to be driven by pupils at the top end of the grade distribution having less growth potential than those lower down. Such ceiling effects may affect changes in grade gaps between ethnic groups which are positioned differently, on average, within the overall distribution of grades. Most of our ethnic minority groups of focus generally attain more highly than their White British peers in both subjects (see [Table 1](#)). As the relative grades of ethnic minority pupils *improve* in maths in 2020, it seems unlikely that ceiling effects drive the observed ethnicity grade gap changes in maths. However the grades for English are on average higher than those for maths. It is therefore possible that ceiling effects are driving some of our results for English in a way which they are not for maths, despite most of the ethnic minority groups achieving more highly than White British pupils, in 'normal' years, in both subjects. Mean grades broken down by subject, group, and year, can be found in [Table B.5](#).

Following [Murphy and Wyness \(2020\)](#), we check the contribution of ceiling effects to the observed ethnicity grade gap changes by partitioning the sample into three equally-sized groups according to pupils' prior attainment at age 11 in the same subject. We then repeat the decomposition of the ethnicity grade gap changes for each of these three groups. The results are shown in [Table 4](#). If ceiling effects were driving our results, we would expect to see only the highest attaining ethnic minority pupils receiving lower relative grades in 2020, and potentially more so in English than in maths. Contrary to this, [Table 4](#) shows the raw, explained, and unexplained components of the grade gap changes for each of the three attainment groups, and reveals that the patterns presented above remain across them. This shows that there remain unexplained factors, such as stereotyping effects, which are contributing to the observed ethnicity grade gap changes throughout the distributions of grades, and suggests ceiling effects are not driving our results.

8. Robustness checks

We check that the decomposition results are robust to multiple alternative specifications. For simplicity, [Table 5](#) shows just the raw, explained, and unexplained components from the decomposition results from each of these sensitivity analyses. In Panel A, we account for possible nonlinearities in the association between pupil and school characteristics with outcomes. The results are consistent with those presented above. In Panel B, we expand our decomposition reference group from the year 2019 alone to a combined 2018 and 2019 group.¹⁹

¹⁹ It is not possible to include 2017 in the reference group due to the need to include lagged school value-added measures in the decompositions. This is

Table 4
Gelbach decomposition by ethnic and prior attainment groups.

	Maths				English			
	P&B	I	BA	BC	P&B	I	BA	BC
Panel A: Low prior attaining								
Raw gap change	0.088*** (0.026)	0.123*** (0.040)	0.177*** (0.028)	0.229*** (0.037)	-0.036 (0.024)	-0.051 (0.038)	-0.115*** (0.030)	0.101** (0.041)
Explained gap change	0.065*** (0.022)	0.003 (0.026)	0.081*** (0.022)	0.078*** (0.025)	0.073*** (0.021)	-0.001 (0.024)	0.057** (0.023)	0.055** (0.025)
Unexplained gap change	0.023 (0.026)	0.119*** (0.036)	0.097*** (0.027)	0.151*** (0.032)	-0.109*** (0.028)	-0.050 (0.038)	-0.171*** (0.029)	0.046 (0.039)
N	312,268				319,447			
Panel B: Medium prior attaining								
Raw gap change	0.145*** (0.025)	0.181*** (0.036)	0.185*** (0.030)	0.129*** (0.046)	-0.041 (0.026)	-0.091** (0.035)	-0.097*** (0.033)	0.041 (0.049)
Explained gap change	0.062*** (0.021)	0.040* (0.022)	0.037* (0.020)	0.015 (0.024)	0.093*** (0.023)	-0.010 (0.024)	0.061*** (0.023)	0.091*** (0.027)
Unexplained gap change	0.083*** (0.028)	0.141*** (0.036)	0.148*** (0.030)	0.115*** (0.044)	-0.134*** (0.030)	-0.081** (0.038)	-0.158*** (0.033)	-0.050 (0.046)
N	313,868				318,017			
Panel C: High prior attaining								
Raw gap change	0.059** (0.027)	0.075*** (0.028)	0.049 (0.033)	0.124** (0.063)	-0.079*** (0.029)	-0.087*** (0.034)	-0.127*** (0.037)	0.046 (0.070)
Explained gap change	0.011 (0.022)	-0.019 (0.021)	-0.012 (0.023)	-0.078** (0.034)	0.085*** (0.024)	-0.014 (0.025)	0.047* (0.025)	0.076** (0.034)
Unexplained gap change	0.048* (0.028)	0.094*** (0.027)	0.061* (0.032)	0.203*** (0.056)	-0.164*** (0.032)	-0.073** (0.033)	-0.173*** (0.037)	-0.031 (0.063)
N	308,454				297,126			

Notes: National Pupil Database. 'P&B' stands for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. The gap is the gap in average grades between the ethnic minority group and White British pupils. Explanatory characteristics are the same as those in Table 2. Standard errors in parentheses, clustered at the school level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

The results are also in line with those in Table 2. In fact, when 2018 is included in the reference group, the estimated negative unexplained contributions to the English grade gap changes become even larger. Moreover, a negative unexplained contribution, statistically significant at the five percent level, is also estimated for Black Caribbean pupils in English.

In Panel C we explore how our results change when we use a dichotomous pass or fail outcome (a pass being a grade 4 and above), rather than a continuous grade, to take into account that grades are ordinal measures of attainment. The grade gap changes and unexplained contributions are smaller when we use this binary outcome, but interestingly the positive and negative unexplained components for maths and English remain. An exception is the raw ethnicity grade gap change for Indian pupils, which becomes statistically indistinguishable from zero. This is most likely because Indian pupils are a particularly high-attaining group and few are observed at this threshold. Finally, in Panel D we restrict our sample to a balanced panel of schools across 2019 and 2020 to explore whether changes in the inclusion of entire schools across years are driving our results. We find that the results are robust to this further restriction.

We also examine the extent to which our results are sensitive to the choice of the English outcome used. So far we have used the higher grade that pupils achieved in English Literature and English Language, as this is the outcome used for performance benchmarks for both schools and pupils (Department for Education, 2020b). Instead, Table B.6 reports separate results for English Literature and English Language. As in Table 2, for both outcomes the point estimates for the explained gap changes are positive and consistently significant (apart from for Indian pupils), indicating that we would expect to see positive grade gap changes for all ethnic minority groups in both English outcomes, given the changes in the groups' characteristics and the returns to those characteristics in 2020 compared to 2019. The

not possible for 2017 given changes in national assessments implemented that year.

unexplained contributions are generally negative for both outcomes, indicating that there remain unexplained factors, such as stereotyping effects, which are leading to the observed ethnicity grade gap changes by reverting the positive impact of the observed characteristics. As the raw gap changes are generally null in the case of English Language and negative in the case of English Literature, these unexplained negative components are considerably larger (about twice as large) in the case of English Literature.

To explore why potential stereotyping effects might be greater for English Literature than English Language, we analyse the correlations between English Literature and English Language grades across the ethnic minority groups. We find the correlation coefficient for White British pupils (0.843) is larger compared to the ethnic minority groups (0.825 for Pakistani and Bangladeshi pupils, 0.815 for Indian pupils, 0.809 for Black African pupils, and 0.816 for Black Caribbean pupils). As shown in Table B.5, this difference appears to be the result of ethnic minority pupils tending to achieve higher, relative to White British pupils, in English Literature than in English Language.²⁰ We therefore report, in Table 6, the raw, explained, and unexplained components of a decomposition using the mean English grade as the outcome, rather than the higher of English Literature and English Language. Though the estimates are smaller than those previously reported, the pattern of the results is unchanged.

²⁰ Establishing why there are differences in ethnic minority pupils' relative grades for English Literature and English Language is beyond the scope of this paper. However, English Language is a more skills-based exam, requiring pupils to demonstrate proficiency in comprehension, analysis, and free writing. English Literature is more knowledge-based and requires pupils to write essays about selected texts. Neither has a coursework element during our years of focus. It may be, then, that the skills assessed in English Literature are different to those assessed in English Language in a way that matters for ethnic minority pupils' performance. For example the English Language syllabi may be more culturally biased than those for English Literature (Ofqual, 2022).

Table 5
Alternative specifications.

	Maths				English			
	P&B	I	BA	BC	P&B	I	BA	BC
Panel A: Non-linear specification								
Raw gap change	0.101*** (0.023)	0.091*** (0.028)	0.115*** (0.024)	0.201*** (0.035)	-0.060*** (0.020)	-0.041 (0.028)	-0.103*** (0.023)	0.082** (0.036)
Explained gap change	0.062*** (0.022)	-0.011 (0.025)	0.006 (0.021)	0.054* (0.030)	0.082*** (0.019)	0.055** (0.022)	0.067*** (0.020)	0.092*** (0.024)
Unexplained gap change	0.039** (0.019)	0.102*** (0.020)	0.109*** (0.019)	0.147*** (0.026)	-0.142*** (0.020)	-0.096*** (0.025)	-0.170*** (0.021)	-0.010 (0.030)
N	934,590				934,590			
Panel B: Combined 2018 and 2019 comparison year								
Raw gap change	0.140*** (0.022)	0.144*** (0.025)	0.105*** (0.021)	0.189*** (0.030)	-0.038** (0.019)	-0.025 (0.024)	-0.118*** (0.021)	0.052 (0.031)
Explained gap change	0.102*** (0.020)	0.034 (0.022)	0.020 (0.019)	0.060** (0.026)	0.103*** (0.018)	0.068*** (0.020)	0.084*** (0.018)	0.111*** (0.022)
Unexplained gap change	0.039** (0.017)	0.110*** (0.018)	0.084*** (0.017)	0.128*** (0.023)	-0.140*** (0.018)	-0.093*** (0.021)	-0.202*** (0.018)	-0.059** (0.027)
N	1,373,869				1,373,869			
Panel C: Binary outcome (grade 4 and above)								
Raw gap change	0.021*** (0.004)	-0.007 (0.004)	0.029*** (0.005)	0.061*** (0.009)	-0.010** (0.004)	-0.022*** (0.004)	-0.013*** (0.004)	0.029*** (0.008)
Explained gap change	0.013*** (0.004)	-0.013*** (0.004)	0.008** (0.004)	0.024*** (0.005)	0.012*** (0.003)	-0.005 (0.003)	0.013*** (0.003)	0.024*** (0.004)
Unexplained gap change	0.008* (0.004)	0.006 (0.004)	0.021*** (0.005)	0.037*** (0.008)	-0.022*** (0.004)	-0.017*** (0.004)	-0.026*** (0.004)	0.005 (0.007)
N	934,590				934,590			
Panel D: Balanced panel of schools								
Raw gap change	0.089*** (0.020)	0.092*** (0.028)	0.107*** (0.023)	0.194*** (0.035)	-0.075*** (0.019)	-0.037 (0.027)	-0.110*** (0.022)	0.077** (0.036)
Explained gap change	0.048** (0.020)	-0.010 (0.023)	0.011 (0.021)	0.044 (0.030)	0.061*** (0.018)	0.039* (0.021)	0.068*** (0.019)	0.091*** (0.024)
Unexplained gap change	0.040*** (0.019)	0.101*** (0.021)	0.096*** (0.019)	0.150*** (0.027)	-0.137*** (0.020)	-0.076*** (0.025)	-0.178*** (0.021)	-0.014 (0.031)
N	891,899				891,899			

Notes: National Pupil Database. 'P&B' stands for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. The reference group is White British. Explanatory characteristics are the same as those in Table 2. Standard errors in parentheses, clustered at the school-level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6
Decomposition using mean English grade.

	P&B	I	BA	BC
Raw gap change	-0.051*** (0.019)	-0.018 (0.028)	-0.078*** (0.022)	0.085** (0.034)
Explained gap change	0.079*** (0.019)	0.054** (0.021)	0.078*** (0.019)	0.098*** (0.024)
Unexplained gap change	-0.130*** (0.019)	-0.073*** (0.023)	-0.156*** (0.020)	-0.013 (0.028)
N	934,590			

Notes: National Pupil Database. 'P&B' stands for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. The reference group is White British. Explanatory characteristics are the same as those in Table 2. Standard errors in parentheses, clustered at the school-level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

9. Discussion and conclusions

An important question in education policy is whether pupils with different characteristics receive systematically different grades in teacher assessments compared to blindly-marked examinations. We contribute new evidence about this question by exploiting a change in assessment methods to examine teachers' predictions of pupils' examination performance. In doing so we use a different comparison to that used in much existing research on teacher assessment, which relies on comparing "blind" (examination) assessments to "non-blind" (teacher) assessments which are intended to consider pupils' written, practical, and oral classwork over an entire academic year. Our approach provides a level of comparability in outcome – in our case examination performance – which is arguably greater than in some of these existing studies.

We find that, when grades are assigned by teachers rather than through externally marked examinations, ethnic minority pupils tend to do relatively better than White British pupils in maths and relatively worse than White British pupils in English. These ethnicity grade gap changes do not appear to be driven by differences across groups and cohorts, time trends, ceiling effects, or by changes across years in the observed characteristics of pupils and schools, or their returns. In fact, while for maths the changes in the levels of and returns to observed characteristics do explain some – but not all – of the observed ethnicity grade gap changes, in the case of English our results suggest that had those changes not occurred, the relative drop in the English performance of ethnic minority pupils that followed the switch to teachers' assessment would have been even larger – roughly double – in magnitude.

There are other potential explanations for the ethnicity grade gap changes that we document. The timing of the announcement to cancel examinations in 2020 allows us to rule out that changes in teaching practices are driving our results, as this announcement came just two months before the examinations were due to begin. Similarly, as teachers were asked to predict precisely how they believed their students would perform had the examinations gone ahead, primarily based on mock examinations completed prior to school closure, we do not believe that our findings are driven by systematic differences in how pupils perform in class-based work and examinations, or by differences in their experiences during school closure.

One limitation of our analysis is that we cannot fully check whether the 2019 and the 2020 cohorts differ in unobserved characteristics. As most time-invariant unobservable characteristics should be captured by the measures of past attainment (at age 11) that we use in our decomposition, any effect of time-varying unobserved characteristics will be included in the unexplained part. However, in order to explain the residual grade gap changes we observe, the effect of unobservables would have to be very large in maths, and very large and working in the opposite direction to the effect of observable characteristics in English. Based on this, we argue that group- and subject-specific teacher stereotyping contributes to at least some part of the ethnicity grade gap changes we document.

Unlike much of the literature on stereotyping, we find both positive and negative unexplained components for a gap in outcomes for the same group within the same broad competency — academic achievement. Our results suggest that teachers hold positive stereotypes about the performance of ethnic minority pupils in maths in high-stakes examinations taken at age 16, and negative stereotypes about the performance of ethnic minority pupils in English. These findings are in line with, and add to, existing research about teachers’ assessments of pupils aged seven to 14 in England (Campbell, 2015; Gibbons and Chevalier, 2008). However they do not conform to common economic approaches whereby stereotypes are held to be manifestations of statistical discrimination (Phelps, 1972) or representativeness heuristics (Bordalo et al., 2016; Esponda et al., 2023), as most ethnic minority groups tend to perform better than White British pupils in both subjects in ‘normal’ years. It may be that teachers’ stereotypes are based on information about pupils at a younger age; for example at school entry, when ethnic minority pupils tend to perform relatively worse in literacy and English than in maths compared to White British pupils (Dustmann et al., 2010). Alternatively they may be informed by the educational choices of older pupils; conditional on prior attainment and family educational background, most ethnic minority pupils are more likely to choose to

study STEM subjects at A Level (ages 16 to 18) than White British pupils (Codioli McMaster, 2017).

Our finding that teacher assigned grades are likely affected at least in part by stereotypes – both positive and negative – is relevant to important policy considerations. Most countries use non-blind assessments at some stages of children’s educational trajectories, and in England, teacher assessments are still considered the primary contingency plan for any future examination disruption (Ofqual, 2023). There is also an ongoing policy debate in England about whether teacher assigned grades should replace high-stakes national examinations because they may be better for pupil well-being, curriculum breadth, and assessing pupils according to competencies that examinations cannot measure (Council of Skills Advisors, 2022). Our results suggest that the benefits of teacher assessments must be weighed against the risk that favourable or unfavourable non-blind assessments reflecting stereotypes might affect pupils’ opportunities, future performance, and education preferences (Burgess et al., 2022; De Benedetto and De Paola, 2023). If teachers’ assessments and expectations differ systematically according to pupils’ characteristics then this could affect gaps in educational outcomes and returns both during schooling and in higher education. With better knowledge of the mechanisms that lead teachers to adopt stereotypes, interventions such as information campaigns could be devised to lessen their effects and make teacher assessments more accurate.

CRedit authorship contribution statement

Hester Burn: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Laura Fumagalli:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Birgitta Rabe:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Data availability

The authors do not have permission to share data.

Appendix A

See Fig. A.1.

Appendix B

See Tables B.1–B.6

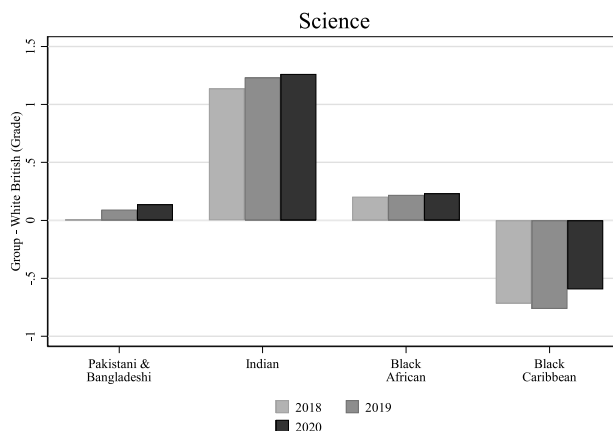


Fig. A.1. Science grade gaps by ethnic group and year.

Notes: National Pupil Database. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The Science grade is the mean of all Science grades awarded (‘double award’ or ‘triple award’). Grade gaps for 2017 are not included as reformed science GCSE qualifications were first awarded in 2018.

Table B.1
Cohort characteristics by year.

	2017	2018	2019	2020
Pupil characteristics				
Male	0.498 (0.500)	0.500 (0.500)	0.500 (0.500)	0.502 (0.500)
Ethnicity:				
White British	0.748 (0.434)	0.732 (0.443)	0.719 (0.450)	0.711 (0.453)
Pakistani & Bangladeshi	0.056 (0.230)	0.061 (0.240)	0.062 (0.242)	0.060 (0.238)
Indian	0.025 (0.155)	0.027 (0.161)	0.027 (0.163)	0.027 (0.162)
Black African	0.029 (0.169)	0.031 (0.175)	0.033 (0.179)	0.035 (0.183)
Black Caribbean	0.013 (0.114)	0.013 (0.115)	0.014 (0.117)	0.013 (0.114)
Multiethnic	0.044 (0.205)	0.046 (0.210)	0.049 (0.217)	0.052 (0.222)
Any Other Ethnic Group	0.085 (0.278)	0.088 (0.284)	0.095 (0.293)	0.102 (0.302)
Free school meals	0.253 (0.435)	0.249 (0.432)	0.246 (0.430)	0.244 (0.430)
Neighbourhood deprivation	0.192 (0.138)	0.192 (0.138)	0.194 (0.139)	0.196 (0.139)
Special educational needs	0.109 (0.311)	0.109 (0.312)	0.113 (0.317)	0.120 (0.325)
First language not English	0.131 (0.337)	0.138 (0.345)	0.141 (0.348)	0.141 (0.349)
Maths prior attainment (age 11)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
English prior attainment (age 11)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)	0.000 (1.000)
School characteristics				
Selective admissions	0.044 (0.206)	0.048 (0.213)	0.047 (0.212)	0.046 (0.209)
Region:				
London	0.138 (0.345)	0.142 (0.349)	0.144 (0.351)	0.143 (0.351)
East of England	0.117 (0.321)	0.116 (0.321)	0.118 (0.322)	0.115 (0.319)
North East	0.150 (0.357)	0.150 (0.357)	0.148 (0.355)	0.148 (0.355)
North West	0.141 (0.348)	0.138 (0.345)	0.142 (0.349)	0.139 (0.346)
South East	0.159 (0.366)	0.161 (0.367)	0.165 (0.371)	0.162 (0.368)
South West	0.098 (0.298)	0.095 (0.293)	0.083 (0.276)	0.092 (0.289)
East Midlands	0.084 (0.277)	0.085 (0.279)	0.087 (0.281)	0.087 (0.282)
West Midlands	0.112 (0.315)	0.113 (0.317)	0.113 (0.317)	0.114 (0.318)
Urban	0.876 (0.330)	0.877 (0.328)	0.877 (0.329)	0.874 (0.332)
Size of cohort	190 (62.5)	189 (63.6)	192 (62.3)	196 (63.0)
School governance type:				
Local Authority Maintained	0.337 (0.473)	0.301 (0.459)	0.270 (0.444)	0.244 (0.429)
Single Academy Trust	0.185 (0.388)	0.192 (0.394)	0.198 (0.398)	0.192 (0.394)
Multi Academy Trust	0.416 (0.493)	0.454 (0.498)	0.502 (0.500)	0.541 (0.498)
Other	0.062 (0.241)	0.054 (0.225)	0.030 (0.171)	0.023 (0.151)
Proportion free school meals	0.252 (0.151)	0.246 (0.148)	0.243 (0.144)	0.240 (0.140)
Average neighbourhood deprivation	0.193 (0.089)	0.193 (0.088)	0.194 (0.089)	0.196 (0.089)

(continued on next page)

Table B.1 (continued).

School value-added (lagged)	0.022 (0.335)	0.033 (0.406)	0.036 (0.423)	0.026 (0.435)
N	449,673	439,279	444,842	489,748

Notes: National Pupil Database. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Free school meals indicates if a pupil is known to have been eligible for free school meals in the past six years. Neighbourhood deprivation is a rank based on the proportion of all children aged 0 to 15 living in income deprived households in the pupil's local area of residence (IDACI score). Prior attainment scores are standardised by year and subject with mean zero and standard deviation of one. School value-added refers to how many average grades higher or lower the pupils in that school achieve across eight qualifying GCSE subjects compared to pupils across the country who score comparatively at age 11. Standard deviations in parentheses.

Table B.2

Cohort characteristics by group.

	WB	PB	I	BA	BC
Pupil characteristics					
Male	0.501 (0.500)	0.496 (0.500)	0.509 (0.500)	0.483 (0.500)	0.483 (0.500)
Free school meals	0.213 (0.410)	0.387 (0.487)	0.144 (0.351)	0.494 (0.500)	0.466 (0.499)
Neighbourhood deprivation	0.172 (0.133)	0.274 (0.115)	0.193 (0.112)	0.316 (0.126)	0.301 (0.123)
Special educational needs	0.118 (0.323)	0.093 (0.291)	0.060 (0.238)	0.097 (0.296)	0.171 (0.377)
First language not English	0.004 (0.062)	0.702 (0.457)	0.595 (0.491)	0.529 (0.499)	0.025 (0.156)
Maths prior attainment (age 11)	0.006 (0.992)	-0.138 (1.035)	0.288 (0.945)	-0.099 (0.998)	-0.330 (1.013)
English prior attainment (age 11)	0.040 (0.991)	-0.266 (0.992)	0.095 (0.965)	-0.110 (0.972)	-0.235 (0.958)
School characteristics					
Selective admissions	0.041 (0.198)	0.039 (0.194)	0.151 (0.358)	0.050 (0.217)	0.013 (0.115)
Region:					
London	0.056 (0.230)	0.262 (0.440)	0.291 (0.454)	0.600 (0.490)	0.640 (0.480)
East of England	0.126 (0.332)	0.070 (0.255)	0.064 (0.245)	0.075 (0.263)	0.055 (0.229)
North East	0.169 (0.375)	0.185 (0.388)	0.063 (0.243)	0.043 (0.203)	0.029 (0.167)
North West	0.157 (0.363)	0.153 (0.360)	0.097 (0.295)	0.055 (0.228)	0.029 (0.169)
South East	0.178 (0.382)	0.079 (0.270)	0.131 (0.337)	0.086 (0.280)	0.043 (0.203)
South West	0.112 (0.315)	0.010 (0.101)	0.028 (0.164)	0.020 (0.140)	0.018 (0.132)
East Midlands	0.095 (0.293)	0.039 (0.192)	0.133 (0.339)	0.042 (0.201)	0.036 (0.186)
West Midlands	0.107 (0.309)	0.202 (0.402)	0.194 (0.395)	0.079 (0.270)	0.149 (0.357)
Urban	0.843 (0.364)	0.988 (0.108)	0.977 (0.151)	0.985 (0.121)	0.989 (0.105)
Size of cohort	192 (64.0)	199 (58.6)	199 (69.4)	185 (54.5)	180 (54.4)

(continued on next page)

Table B.2 (continued).

School governance type:					
Local Authority Maintained	0.274 (0.446)	0.362 (0.481)	0.264 (0.441)	0.356 (0.479)	0.345 (0.475)
Single Academy Trust	0.188 (0.391)	0.153 (0.360)	0.254 (0.436)	0.208 (0.406)	0.192 (0.394)
Multi Academy Trust	0.491 (0.500)	0.462 (0.499)	0.459 (0.498)	0.409 (0.492)	0.431 (0.495)
Other	0.046 (0.209)	0.023 (0.150)	0.023 (0.150)	0.028 (0.165)	0.032 (0.176)
Proportion free school meals	0.221 (0.128)	0.375 (0.162)	0.242 (0.146)	0.356 (0.171)	0.377 (0.161)
Average neighbourhood deprivation	0.178 (0.083)	0.265 (0.082)	0.206 (0.075)	0.269 (0.082)	0.275 (0.077)
School value-added (lagged)	-0.019 (0.381)	0.164 (0.414)	0.299 (0.454)	0.181 (0.420)	0.093 (0.426)
N	1,326,331	109,634	48,111	58,848	24,388

Notes: National Pupil Database. 'WB' stands for White British, 'P&B' stands for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England, with no missing data, in years 2017, 2018, 2019 or 2020. Free school meals indicates if a pupil is known to have been eligible for free school meals in the past six years. Neighbourhood deprivation is a rank based on the proportion of all children aged 0 to 15 living in income deprived households in the pupil's local area of residence (IDACI score). Prior attainment scores are standardised by year and subject with mean zero and standard deviation of one. School value-added refers to how many average grades higher or lower the pupils in that school achieve across eight qualifying GCSE subjects compared to pupils across the country who score comparatively at age 11. Standard deviations in parentheses.

Table B.3
Detailed Gelbach decomposition of the change in grade gaps by ethnic group.

	Maths				English			
	P&B	I	BA	BC	P&B	I	BA	BC
Raw gap change	0.101*** (0.023)	0.091*** (0.028)	0.115*** (0.024)	0.201*** (0.035)	-0.060*** (0.020)	-0.041 (0.028)	-0.103*** (0.023)	0.082** (0.036)
Explained gap change	0.065*** (0.022)	-0.009 (0.024)	0.018 (0.021)	0.056* (0.030)	0.080*** (0.019)	0.036* (0.022)	0.075*** (0.020)	0.104*** (0.025)
<i>Amount explained by:</i>								
Male	0.001 (0.002)	0.000 (0.002)	0.002 (0.002)	0.003* (0.002)	0.003 (0.003)	0.004 (0.004)	-0.004 (0.004)	0.005 (0.005)
Socio-economic status:								
Free school meals	-0.002 (0.002)	0.009*** (0.002)	0.002 (0.003)	-0.001 (0.004)	0.003 (0.002)	0.007*** (0.002)	0.012*** (0.003)	0.008* (0.005)
Neighbourhood deprivation	0.008*** (0.003)	0.003 (0.002)	0.014*** (0.004)	0.007 (0.004)	0.012*** (0.003)	0.004* (0.002)	0.019*** (0.004)	0.011** (0.005)
Special educational needs	0.004*** (0.001)	0.007*** (0.001)	0.004*** (0.001)	-0.003 (0.002)	0.005*** (0.002)	0.007*** (0.002)	0.006*** (0.002)	0.000 (0.004)
First language not English	0.011 (0.009)	0.008 (0.008)	0.005 (0.007)	0.000 (0.001)	0.011 (0.010)	0.009 (0.008)	0.006 (0.007)	0.000 (0.001)
Prior attainment (age 11)	0.024 (0.016)	-0.026 (0.019)	-0.015 (0.016)	0.027 (0.026)	0.002 (0.010)	0.036*** (0.014)	0.030*** (0.012)	0.047*** (0.016)
School characteristics:								
Selective admissions	0.000 (0.001)	-0.002 (0.004)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	-0.005 (0.004)	0.001 (0.001)	0.002 (0.001)
Region	0.007 (0.005)	-0.001 (0.005)	0.013 (0.009)	0.014 (0.011)	-0.012** (0.006)	-0.019*** (0.006)	-0.025** (0.011)	-0.024** (0.012)
Urban	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)	-0.001 (0.003)
Size of cohort	-0.003* (0.001)	-0.002 (0.002)	0.003** (0.001)	0.006*** (0.002)	-0.002 (0.001)	-0.002 (0.002)	0.002** (0.001)	0.005*** (0.002)
School governance type	-0.001 (0.001)	0.000 (0.001)	0.001 (0.001)	0.001 (0.001)	0.000 (0.002)	0.001 (0.001)	0.002 (0.001)	0.001 (0.001)
Prop. free school meals	0.019 (0.015)	0.001 (0.002)	0.016 (0.014)	0.020 (0.016)	0.048*** (0.016)	0.006** (0.003)	0.043*** (0.015)	0.049*** (0.017)
Avg. neigh'hood deprivation	-0.023 (0.014)	-0.006 (0.004)	-0.024 (0.014)	-0.026* (0.016)	-0.003 (0.016)	-0.003 (0.005)	-0.003 (0.016)	-0.002 (0.018)
School value-added (lagged)	0.015* (0.009)	-0.003 (0.009)	-0.008 (0.008)	0.004 (0.010)	0.012 (0.009)	-0.008 (0.010)	-0.011 (0.008)	0.003 (0.010)
Unexplained gap change	0.036* (0.019)	0.100*** (0.020)	0.096*** (0.019)	0.145*** (0.026)	-0.140*** (0.020)	-0.077*** (0.025)	-0.179*** (0.021)	-0.022 (0.030)
N	934,590				934,590			

Notes: National Pupil Database. 'P&B' stands for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. The gap is the gap in average grades between the ethnic minority group and White British pupils. Free school meals indicates if a pupil is known to have been eligible for free school meals in the past six years. Neighbourhood deprivation is a rank based on the proportion of all children aged 0 to 15 living in income deprived households in the pupil's local area of residence (IDACI score). Prior attainment scores are standardised by year and subject with mean zero and standard deviation of one. Region contains 9 categories and school governance type 4 categories. School value-added refers to how many average grades higher or lower the pupils in that school achieve across eight qualifying GCSE subjects compared to pupils across the country who score comparatively at age 11. Standard errors in parentheses, clustered at the school level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.4
Gelbach decomposition returns.

	Maths		English	
	2019	*2020	2019	*2020
Constant	5.057*** (0.033)	0.468*** (0.033)	5.831*** (0.037)	0.134*** (0.038)
Male	-0.089*** (0.006)	-0.130*** (0.007)	-0.610*** (0.006)	0.086*** (0.007)
Free school meals	-0.355*** (0.005)	-0.050*** (0.007)	-0.418*** (0.006)	-0.018** (0.007)
Neighbourhood deprivation	-1.115*** (0.023)	0.079*** (0.027)	-1.138*** (0.024)	0.114*** (0.028)
Special educational needs	-0.198*** (0.009)	-0.074*** (0.010)	-0.514*** (0.009)	-0.029*** (0.011)
First language not English	0.170*** (0.011)	0.020 (0.013)	0.195*** (0.012)	0.022 (0.014)
Prior attainment (age 11)	1.412*** (0.004)	-0.056*** (0.003)	0.981*** (0.003)	0.014*** (0.004)
Selective admissions	0.420*** (0.032)	-0.019 (0.025)	0.435*** (0.030)	-0.042 (0.027)
Region (ref: London):				
East of England	-0.022 (0.023)	-0.011 (0.022)	-0.231*** (0.025)	0.074*** (0.025)
North East	0.003 (0.022)	-0.010 (0.021)	-0.156*** (0.024)	0.057** (0.023)
North West	-0.062*** (0.022)	-0.016 (0.022)	-0.135*** (0.026)	0.019 (0.025)
South East	-0.022 (0.022)	-0.032 (0.022)	-0.225*** (0.024)	0.027 (0.024)
South West	0.005 (0.025)	-0.029 (0.025)	-0.237*** (0.027)	0.074** (0.028)
East Midlands	0.009 (0.027)	-0.080*** (0.026)	-0.189*** (0.028)	0.024 (0.030)
West Midlands	-0.030 (0.023)	-0.043* (0.022)	-0.186*** (0.027)	0.022 (0.025)
Urban	0.003 (0.017)	0.021 (0.015)	0.019 (0.017)	-0.008 (0.019)
Size of cohort	-0.000 (0.000)	-0.000*** (0.000)	-0.000 (0.000)	-0.000*** (0.000)
School governance type (ref: Community):				
Single Academy Trust	0.001 (0.016)	0.004 (0.016)	0.022 (0.018)	0.007 (0.018)
Multi Academy Trust	0.025* (0.013)	-0.008 (0.014)	0.044*** (0.014)	-0.013 (0.014)
Other	0.000 (0.034)	0.041 (0.037)	-0.017 (0.039)	0.000 (0.045)
Proportion free school meals	-0.023 (0.100)	0.137 (0.104)	-0.293*** (0.104)	0.322*** (0.113)
Average neighbourhood deprivation	-0.067 (0.156)	-0.262 (0.159)	0.688*** (0.166)	-0.035 (0.181)
School value-added (lagged)	0.615*** (0.016)	-0.035** (0.017)	0.618*** (0.018)	-0.050*** (0.018)
N	934,590		934,590	

Notes: National Pupil Database. Columns labelled '2019' refer to the returns to a characteristic in 2019. Columns labelled '*2020' are the interaction terms for each characteristic and the 2020 indicator. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The English grade is the highest of English Literature and English Language if pupils received grades for both. The gap is the gap in average grades between the ethnic minority group and White British pupils. Free school meals indicates if a pupil is known to have been eligible for free school meals in the past six years. Neighbourhood deprivation is a rank based on the proportion of all children aged 0 to 15 living in income deprived households in the pupil's local area of residence (IDACI score). Prior attainment scores are standardised by year and subject with mean zero and standard deviation of one. School value-added refers to how many average grades higher or lower the pupils in that school achieve across eight qualifying GCSE subjects compared to pupils across the country who score comparatively at age 11. Standard errors in parentheses, clustered at the school level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.5
Mean grades by subject, group, and year.

	WB	PB	I	BA	BC
Maths					
2017	4.651 (2.003)	4.614 (2.045)	5.797 (2.063)	4.804 (1.987)	3.948 (1.883)
2018	4.699 (1.996)	4.655 (2.037)	5.813 (2.046)	4.809 (1.978)	3.829 (1.858)
2019	4.697 (1.995)	4.734 (2.060)	5.920 (2.071)	4.787 (1.999)	3.802 (1.826)
2020	4.974 (1.959)	5.111 (2.011)	6.287 (1.964)	5.177 (1.908)	4.279 (1.784)
N	1,326,331	109,634	48,111	58,848	24,388
English (highest)					
2017	5.146 (1.898)	5.331 (1.813)	6.125 (1.767)	5.558 (1.762)	4.900 (1.786)
2018	5.132 (1.897)	5.296 (1.819)	6.088 (1.766)	5.525 (1.772)	4.761 (1.793)
2019	5.130 (1.899)	5.340 (1.836)	6.118 (1.793)	5.493 (1.776)	4.698 (1.761)
2020	5.326 (1.827)	5.475 (1.754)	6.273 (1.679)	5.585 (1.683)	4.976 (1.646)
N	1,326,331	109,634	48,111	58,848	24,388
English Literature					
2017	4.720 (1.973)	4.986 (1.910)	5.810 (1.856)	5.253 (1.837)	4.550 (1.892)
2018	4.749 (1.971)	5.011 (1.905)	5.817 (1.839)	5.253 (1.850)	4.453 (1.898)
2019	4.790 (1.960)	5.100 (1.900)	5.853 (1.849)	5.232 (1.826)	4.420 (1.829)
2020	5.089 (1.881)	5.292 (1.809)	6.089 (1.724)	5.403 (1.724)	4.764 (1.702)
N	1,268,229	103,987	46,455	57,367	23,776
English Language					
2017	4.758 (1.842)	4.809 (1.730)	5.548 (1.742)	4.942 (1.695)	4.387 (1.682)
2018	4.756 (1.838)	4.783 (1.726)	5.521 (1.740)	4.929 (1.702)	4.245 (1.658)
2019	4.754 (1.840)	4.822 (1.767)	5.593 (1.762)	4.928 (1.725)	4.207 (1.666)
2020	5.133 (1.803)	5.202 (1.724)	5.988 (1.670)	5.288 (1.652)	4.711 (1.605)
N	1,313,822	108,428	47,653	58,209	24,189
English mean					
2017	4.731 (1.826)	4.892 (1.726)	5.676 (1.702)	5.095 (1.663)	4.462 (1.693)
2018	4.742 (1.825)	4.893 (1.727)	5.665 (1.697)	5.089 (1.680)	4.348 (1.691)
2019	4.756 (1.826)	4.945 (1.750)	5.710 (1.719)	5.077 (1.687)	4.306 (1.670)
2020	5.096 (1.807)	5.234 (1.723)	6.031 (1.654)	5.339 (1.646)	4.731 (1.609)
N	1,326,331	109,634	48,111	58,848	24,388

Notes: National Pupil Database. 'WB' stands for White British, 'P&B' for Pakistani and Bangladeshi, 'I' Indian, 'BA' Black African, 'BC' Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. Standard deviations in parentheses. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.6
Gelbach decomposition by English outcome.

	English Language				English Literature			
	P&B	I	BA	BC	P&B	I	BA	BC
Raw gap change	0.001 (0.021)	0.016 (0.028)	-0.020 (0.023)	0.125*** (0.035)	-0.108*** (0.023)	-0.063** (0.030)	-0.128*** (0.024)	0.044 (0.037)
Explained gap change	0.091*** (0.019)	0.066*** (0.021)	0.091*** (0.019)	0.104*** (0.024)	0.066*** (0.022)	0.041* (0.023)	0.070*** (0.021)	0.092*** (0.026)
<i>Amount explained by:</i>								
Male	0.004 (0.003)	0.004 (0.004)	-0.003 (0.004)	0.006 (0.005)	0.004 (0.003)	0.006 (0.005)	-0.005 (0.004)	0.005 (0.006)
Socio-economic status	0.008* (0.004)	0.012*** (0.003)	0.019*** (0.005)	0.008 (0.006)	0.015*** (0.005)	0.011*** (0.004)	0.033*** (0.006)	0.021*** (0.007)
Special educational needs	0.006*** (0.002)	0.010*** (0.002)	0.007*** (0.002)	-0.002 (0.004)	0.004 (0.002)	0.004* (0.002)	0.005** (0.002)	0.001 (0.004)
First language not English	0.027*** (0.009)	0.022*** (0.008)	0.017** (0.007)	0.001 (0.001)	0.010 (0.011)	0.007 (0.009)	0.004 (0.008)	0.000 (0.001)
Prior attainment (age 11)	-0.001 (0.010)	0.037*** (0.014)	0.030*** (0.012)	0.045*** (0.016)	-0.009 (0.011)	0.032** (0.014)	0.026** (0.012)	0.036** (0.016)
School characteristics	0.035*** (0.008)	-0.017** (0.008)	0.030*** (0.011)	0.043*** (0.012)	0.021** (0.009)	-0.023*** (0.008)	0.012 (0.012)	0.020 (0.013)
School value-added (lagged)	0.014* (0.008)	-0.002 (0.009)	-0.008 (0.008)	0.004 (0.009)	0.021* (0.011)	0.005 (0.011)	-0.005 (0.010)	0.008 (0.011)
Unexplained gap change	-0.091*** (0.020)	-0.051** (0.024)	-0.111*** (0.020)	0.021 (0.029)	-0.174*** (0.022)	-0.105*** (0.027)	-0.199*** (0.023)	-0.047 (0.031)
N	924,345				881,399			

Notes: National Pupil Database. ‘P&B’ stands for Pakistani and Bangladeshi, ‘I’ Indian, ‘BA’ Black African, ‘BC’ Black Caribbean. Sample comprises age 16 pupils in mainstream schools in England with no missing data. Exam grades are the May and June (end of year) examinations only. The gap is the gap in average grades between the ethnic minority group and White British pupils. Socio-economic status includes free school meals indicator and rank based on the proportion of all children aged 0 to 15 living in income deprived households in the pupil’s local area of residence (IDACI score). Prior attainment (age 11) includes subject-specific age 11 attainment score, standardised by year with mean zero and standard deviation of one. School characteristics are indicators for selectivity, urban, region (9 categories), and school governance type (4 categories), and continuous measures of cohort size, proportion of pupils eligible for free school meals, and average neighbourhood deprivation (IDACI) score of the school. School value-added refers to how many average grades higher or lower the pupils in that school achieve across eight qualifying GCSE subjects compared to pupils across the country who score comparatively at age 11. Standard errors in parentheses, clustered at the school level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

References

- Alan, S., Ertac, S., Mumcu, I., 2018. Gender stereotypes in the classroom and effects on achievement. *Rev. Econ. Stat.* 100 (5), 876–890.
- Alesina, A., Carlana, M., La Ferrara, E., Pinotti, P., 2018. Revealing Stereotypes: Evidence from Immigrants in Schools. Technical Report, National Bureau of Economic Research.
- Baird, J., 1998. What's in a name? Experiments with blind marking in A-level examinations. *Educ. Res.* 40 (2), 191–202.
- Black, N., de New, S.C., 2020. Short, heavy and underrated? Teacher assessment biases by children's body size. *Oxford Bull. Econ. Stat.* 82 (5), 961–987.
- Blinder, A.S., 1973. Wage discrimination: reduced form and structural estimates. *J. Hum. Resour.* 436–455.
- Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2016. Stereotypes. *Q. J. Econ.* 131 (4), 1753–1794.
- Botelho, F., Madeira, R.A., Rangel, M.A., 2015. Racial discrimination in grading: Evidence from Brazil. *Am. Econ. J. Appl. Econ.* 7 (4), 37–52.
- Burgess, S., Greaves, E., 2013. Test scores, subjective assessment, and stereotyping of ethnic minorities. *J. Labor Econ.* 31 (3), 535–576.
- Burgess, S., Hauberg, D.S., Rangvid, B.S., Sievertsen, H.H., 2022. The importance of external assessments: High school math and gender gaps in STEM degrees. *Econ. Educ. Rev.* 88, 102267.
- Campbell, T., 2015. Stereotyped at seven? Biases in teacher judgement of pupils' ability and attainment. *J. Soc. Policy* 44 (3), 517–547.
- Carlana, M., 2019. Implicit stereotypes: Evidence from teachers' gender bias. *Q. J. Econ.* 134 (3), 1163–1224.
- Codioli McMaster, N., 2017. Who studies STEM subjects at a level and degree in England? An investigation into the intersections between students' family background, gender and ethnicity in determining choice. *Br. Educ. Res. J.* 43 (3), 528–553.
- Council of Skills Advisors, 2022. In: Council of Skills Advisors (Ed.), *Learning and Skills for Economic Recovery, Social Cohesion and a More Equal Britain*. Technical Report, Labour Party.
- De Benedetto, M.A., De Paola, M., 2023. Immigration and teacher bias towards students with an immigrant background. *Econ. Policy* 38 (113), 107–154.
- Dee, T.S., 2005. A teacher like me: Does race, ethnicity, or gender matter? *Amer. Econ. Rev.* 95 (2), 158–165.
- Del Bono, E., Fumagalli, L., Holford, A., Rabe, B., 2022. University Access: The Role of Background and Covid-19 Throughout the Application Process. University of Essex, Institute for Social and Economic Research (ISER), ISER Working Paper Series 2022–07.
- Department for Education, 2020a. Ad-Hoc Notice: Early entry into GCSE Examinations in England. Technical Report, Department for Education.
- Department for Education, 2020b. Secondary Accountability Measures: Guide for Maintained Secondary Schools, Academies and Free Schools. Technical Report, Department for Education.
- Department for Education, 2021. GCSE Attainment and Lifetime Earnings: Research Report. Technical Report, Department for Education.
- Department for Education, 2022. Post-Qualification Admissions Consultation Response. Technical Report, Department for Education.
- Dustmann, C., Machin, S., Schönberg, U., 2010. Ethnicity and educational achievement in compulsory schooling. *Econ. J.* 120 (546), F272–F297.
- Eivers, E., Worth, J., Ghosh, A., 2020. Home Learning during COVID-19: Findings from the Understanding Society Longitudinal Study. Technical Report, National Foundation for Educational Research.
- Esponda, I., Oprea, R., Yuksel, S., 2023. Seeing what is representative. *Q. J. Econ.* 2607–2657.
- Fryer, Jr., R.G., 2017. The production of human capital in developed countries: Evidence from 196 randomized field experiments. In: *Handbook of Economic Field Experiments*, vol. 2, Elsevier, pp. 95–322.
- Gelbach, J.B., 2016. When do covariates matter? And which ones, and how much? *J. Labor Econ.* 34 (2), 509–543.
- Gibbons, S., Chevalier, A., 2008. Assessment and age 16+ education participation. *Res. Papers Educ.* 23 (2), 113–123.
- Hanna, R.N., Linden, L.L., 2012. Discrimination in grading. *Am. Econ. J. Econ. Policy* 4 (4), 146–168.
- Joint Council for Qualifications, 2020. GCSE (Full Course) Outcomes for Main Grade Set for Each Jurisdiction (Age 16): Results Summer 2020. Technical Report, Joint Council for Qualifications.
- Lavy, V., 2008. Do gender stereotypes reduce girls' or boys' human capital outcomes? evidence from a natural experiment. *J. Public Econ.* 92 (10–11), 2083–2105.
- Lavy, V., Megalokonomou, R., 2019. Persistency in Teachers' Grading Bias and Effects on Longer-Term Outcomes: University Admissions Exams and Choice of Field of Study. Technical Report, National Bureau of Economic Research.
- Lavy, V., Sand, E., 2018. On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases. *J. Public Econ.* 167, 263–279.
- Lindahl, E., 2007. Comparing Teachers' Assessments and National Test Results: Evidence from Sweden. Technical Report, Working Paper.
- Machin, S., McNally, S., Ruiz-Valenzuela, J., 2020. Entry through the narrow door: The costs of just failing high stakes exams. *J. Public Econ.* 190, 104224.
- Murphy, R., Wyness, G., 2020. Minority report: the impact of predicted grades on university admissions of disadvantaged groups. *Educ. Econ.* 28 (4), 333–350.
- Oaxaca, R., 1973. Male-female wage differentials in urban labor markets. *Internat. Econom. Rev.* 693–709.
- Office for National Statistics, 2022. Population of England and Wales. Technical Report, Office for National Statistics.
- Ofqual, 2020a. Appeals for GCSE, AS, A Level and Project: 2018 to 2019 Academic Year. Technical Report, Office of Qualifications and Examinations Regulation.
- Ofqual, 2020b. Provisional Entries for GCSE, AS and A Level: Summer 2020 Exam Series. Technical Report, Office of Qualifications and Examinations Regulation.
- Ofqual, 2020c. Summer 2020 Grades For GCSE, AS and A level, Extended Project Qualification and Advanced Extension Award in maths: Information for Heads of Centre, Heads of Department/Subject Leads and Teachers on the Submission of Centre Assessment Grades. Updated 22 May 2020. Technical Report, Office of Qualifications and Examinations Regulation.
- Ofqual, 2020d. Student-Level Equalities Analyses for GCSE and A Level: Summer 2020. Technical Report, Office of Qualifications and Examinations Regulation.
- Ofqual, 2021a. Research and Analysis: Centre Judgements: Teaching Staff Survey and Interviews, Summer 2020. Technical Report, Office of Qualifications and Examinations Regulation.
- Ofqual, 2021b. Research and Analysis: Centre Judgements: Teaching Staff Survey, Summer 2020. Technical Report, Office of Qualifications and Examinations Regulation.
- Ofqual, 2022. Consultation Outcome: Guidance on Designing and Developing Accessible Assessments. Technical Report, Office of Qualifications and Examinations Regulation.
- Ofqual, 2023. Guidance on Collecting Evidence of Student Performance to Ensure Resilience in the Qualifications System. Technical Report, Office of Qualifications and Examinations Regulation.
- Phelps, E.S., 1972. The statistical theory of racism and sexism. *Am. Econ. Rev.* 62 (4), 659–661.
- Starck, J.G., Riddle, T., Sinclair, S., Warikoo, N., 2020. Teachers are people too: Examining the racial bias of teachers compared to other American adults. *Educ. Res.* 49 (4), 273–284.
- Tanjitpiyanond, P., Jetten, J., Peters, K., 2022. How economic inequality shapes social class stereotyping. *J. Exp. Soc. Psychol.* 98, 104248.
- Terrier, C., 2020. Boys lag behind: How teachers' gender biases affect student achievement. *Econ. Educ. Rev.* 77, 101981.
- Van Ewijk, R., 2011. Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Econ. Educ. Rev.* 30 (5), 1045–1058.