RESEARCH AND ANALYSIS

Awarding GCSE, AS, A level, advanced extension awards and extended project qualifications in summer 2020: interim report

ofqual

Contents

1 Executive summary

The 2020 exam series has been exceptional. Early in 2020, when coronavirus (COVID-19) was first reported, we modified our normal contingency planning with exam boards to include widespread staff absences and/or large numbers of students being unable to take one or more of their exams due to illness or self-isolation.

On 18 March 2020, the Secretary of State announced[1] that the summer 2020 exam series would be cancelled in order to help fight the spread of coronavirus (COVID-19) and that students due to sit the exams would be awarded a grade based on an assessment of the grade they would have been most likely to achieve had exams gone ahead. On 23 March, in a written statement to Parliament,[2] the Secretary of State explained the government's intention that results would be issued to this year's cohort based on a range of evidence and data, including performance on mock exams and non-exam assessment. This would be achieved by exam boards producing calculated grades ensuring the distribution of grades follows a similar pattern to that in other years, so that this year's students do not face a systemic disadvantage as a consequence of circumstances this year.

Since then, we have been working with exam boards to enable the award of grades for GCSE, AS, A level, Extended Project Qualification and Advanced Extension Award this summer, so that students can move on to sixth form, college, higher education, training, apprenticeships or employment.

Our aim in this work was to use an approach that was, as far as possible, fair to students who had been unable to sit their exams this summer. That approach should also ensure that the results issued had a similar value to grades issued in any other year, so that those using them to select students (sixth forms, universities, employers, etc.) could have confidence that their worth was in line with previous years. A critical factor in achieving that was maintaining overall national standards relative to previous years.

We considered many different options, but it was apparent that the best judges of the relative ability of students in a school or college were the teachers who had been preparing these students for their exams, tracking their progress relative to target grades, and, in the case of A level students applying to higher education, providing estimated grades.

We therefore asked teachers to provide, for each student for each subject they were entered for, a centre assessment grade (CAG) which represented the grade that student would have been most likely to achieve if teaching and learning had continued and students had taken their exams as planned.

We also asked teachers to provide a rank order of students for each grade for each subject. There were several reasons for this. First, we know from research evidence that people are better at making relative judgements than absolute judgements and that teachers' judgements tend to be more accurate when they are ranking students rather than estimating their future attainment. The research literature suggests that, in estimating the grades students are likely to achieve, teachers tend to be optimistic (although not in all cases). That is not surprising, teachers want to do the best for

1 Schools, colleges and early years settings to close.
2 Impact of Covid-19 on Summer Exams: Written statement - HLWS170

their students, and the analysis we carried out immediately after CAGs were submitted bears this out.

Exam boards adapted their IT portals to enable schools and colleges to submit over 5 million CAGs and rank order positions in early June 2020. This was achieved despite teachers often working remotely with limited access to school premises, and with exam board technical teams also working remotely.

Our initial analysis of the CAGs showed that they were, in general, optimistic (although not always) and the combined effect would be likely to lead to overall national results that were implausibly high. If we had awarded grades based on CAGs we would have seen overall results increase by far more than we have ever seen in a single year. At A level[3], we would have seen the percentage of A* grades go up by 6 percentage points from 7.7% of grades in 2019 to 13.9% of grades this year, and the percentage of grades that were B and above increase by over 13 percentage points from 51.1% in 2019 to 65% this year.

This optimism was not surprising. It is what is suggested in the research literature and data published every year by UCAS – schools and colleges tend to be optimistic when estimating the grades that students are likely to achieve. Our interviews with teachers, after CAGs had been submitted, confirmed this trend. Almost all the teachers we interviewed told us that they had generally predicted how the students would perform on a ‘good day’. Although they knew that every year some students underperform or have a bad day, this was not the basis of their judgements. This might be as expected, but the cumulative effect of this optimism, if reflected in the final results, would have undermined confidence in those results.

Standardisation was not solely implemented to ensure that grades were not, overall, excessively high this year. The key purpose was to ensure fairness to students within the 2020 cohort. Without standardisation there was the potential for students to be unfairly advantaged or disadvantaged, depending on the school or college they attended and the approach they took. A key motivation for the design of the approach to standardisation that we took was to remove this potential inequality and, as far as possible, ensure that a grade represents the same standard, irrespective of the school or college they attended.

Given the circumstances this summer, we had no opportunity to put in place a system of national standardisation, to guide teachers in making their judgements consistently across the country. It was, though, essential that we put in place a mechanism to standardise those judgements being made in many thousands of different schools and colleges, in the interests of fairness to students. It was likely that different schools and colleges would take different approaches to generating CAGs and rank orders and that was likely to generate different levels of optimism in different centres (or perhaps in different subjects). That is what we saw.

It was important, then, to have a system to standardise teachers’ judgements across schools and colleges, and that all exam boards took the same approach. A parallel can be drawn here with moderation of teachers’ marking of coursework (non-exam assessment) in a normal year. Even when exam boards provide marking criteria, training materials and events to guide teachers in their marking, exam boards

[3] Note that 2019 figures are for England only as reported by JCQ.

moderate the marking to ensure a common standard is applied, in the interests of fairness to students overall.

The rank order data allowed us to make fine-grained adjustments which is fairest for students. For example, with this information we could prioritise students for an upwards or downwards adjustment where necessary. Without the facility to perform this fine-grained adjustment it may have been necessary to adjust larger groups of students leading to an over or under adjustment.

In April and May we worked with technical experts across the sector to test a range of different statistical standardisation models using data from previous years. In selecting the final model, we chose the one that most accurately predicted students' grades in a way that did not systematically affect groups of students with particular protected characteristics. We also considered operational issues – how easy it was to implement the approaches consistently across all four exam boards – and transparency – how easy it was to explain to schools and colleges how the model worked.

Our preferred model – known as the Direct Centre Performance model (DCP) – works by predicting the distribution of grades for each individual school or college. That prediction is based on the historical performance of the school or college in that subject taking into account any changes in the prior attainment of candidates entering this year compared to previous years. This was fine-tuned to take account of known issues such as centres with small cohorts of students, small-entry subjects, and tiered subjects. Decisions were also made on the number of years of historical data included in the model. The details of these decisions are set out in this report and are formalised in the regulations[4] we put in place for summer 2020.

Where schools and colleges had a relatively small cohort for a subject – fewer than 15 students when looking across the current entry and the historical data – the standardisation model put more weight on the CAGs. Since small teaching groups are more common for AS and A level than for GCSE, and given that the CAGs tended to be optimistic, it means that the outcomes in some AS and A level subjects are much higher this year. However, there is no statistical model that can reliably predict grades for particularly small groups of students. We have therefore used the most reliable evidence available, which is the CAGs.

Overall, A level results in England have increased by 2.4% at grade A and above compared to 2019. This is a larger change than observed in a typical year (for example, there was a 1% decrease in outcomes between 2018 and 2019).

Across all subjects and all centres, 96.4% of final calculated grades are the same as, or within one grade of the CAG submitted. A small percentage were adjusted by 2 grades or more, in some cases because it appeared that the centre's CAGs were very much higher than the historical results in the centre.

In any year, there is measurement uncertainty in the assessment process. This may be due to exam questions being asked that match well with one student's strengths but are poorly matched with the strengths of another student with the same overall level of ability. This year, there is also uncertainty in the results and the challenge is heightened by the absence of any formal assessment information on which to assess students. Based on the testing of the approaches applied this summer using

[4] Requirements for the calculation of results in summer 2020

results data from 2019, 51 of the 55 A levels tested had accurate predictions for more than 90% of students within plus or minus one grade[5]. This figure was lower for GCSE (12 out of 22 subjects) which was likely due to a combination of the grade scaling being longer for GCSE compared to A level and some limitations of the testing. GCSE English language, English literature and mathematics all had above 99% of students receiving results accurate within plus or minus one grade. Overall, the levels of predictive accuracy are broadly comparable to measures of marking consistency across an equivalent range of subjects.

To understand the impact of potential advantage or disadvantage across different demographic and socio-economic groups we have also performed an equalities analysis of calculated grades. The analyses show no evidence that this year's process of awarding grades has introduced bias.

This interim report provides a description of the process for all qualifications and presents analyses of CAGs and calculated grades for AS and A level. Analysis of GCSE CAGs and calculated grades will be published on GCSE results day. A final report will be published later in the year when we have completed our evaluation of this summer's results.

Throughout the development and testing of the model, and in its implementation, we have taken all possible steps to ensure the process is as fair as it can be and, where possible, have taken design decisions in the students' favour. For example, we have used calculations that assume that all students would have attended for all of their assessments this summer. In reality this would not have happened.

We know that, just as in any year, some students will be disappointed with their results. Some students may think that, had they taken their exams, they would have achieved higher grades. We will never know. But for those students who do wish to improve their grades, there will be an autumn exam series.

Where possible, we have urged sixth forms, colleges and universities to be flexible in their selection this year. Overall, the results delivered this year will have met the aim of enabling large numbers of students to move on to the next stages of their lives.

[5] These A level figures are calculated once entries from schools and colleges with fewer than 10 entries in the subject have been removed.

2 Background

2.1 Context

On 18 March 2020 the Secretary of State for Education told Parliament that, in response to the coronavirus (COVID-19) pandemic, schools and colleges in England would shut to all but the children of critical workers and vulnerable children after 20 March. In line with these measures, exams scheduled for the summer would not take place.[6] The Secretary of State said that the government would work with the education sector and with Ofqual to make sure students who were preparing to take GCSEs, AS and A level exams in the summer would not be unfairly penalised.

On 23 March 2020, in a written statement to the House of Commons, the Secretary of State confirmed the government's priority was that students could move to the next stage of their lives and that GCSE, AS and A level students would receive a grade that reflected their work.[7] The statement explained the government's intention that 'a grade will be awarded this summer based on the best available evidence, including any non-exam assessment that students have already completed. There will also be an option, for students who do not feel this grade reflects their performance, to sit an exam at the earliest reasonable opportunity once schools are open again'.

In the direction we received on 31 March 2020[8] it was confirmed that '[i]n order to mitigate the risk to standards as far as possible, the approach should be standardised across centres' and that distribution of grades should follow a similar profile to that in previous years.

Since these announcements we have worked with others from across the sector to develop an approach that enables the fairest possible award of grades in these qualifications, in the absence of any exams.

Our aims in this work were to ensure that students would receive grades to enable them to move on to the next stages of their lives without further disruption; that the grades would have the same currency as those of any other year; and that the approach would be as fair as it could be.

To support this work, we consulted extensively and received unprecedented numbers of responses – with over 12,500 responses to our initial policy consultation[9]. Through our various public consultations, we have sought views from groups that represent students and teachers, from many individual students and parents of students who had expected to take exams this summer, about the way the arrangements might affect them. This information was considered when making decisions regarding the approach to awarding grades this summer. We also brought

6 House of Commons Hansard, 18 March 2020 5:16pm, Column 1083

7 House of Commons Hansard, 23 March 2020, volume 674

8 Direction under S129(6) of the Apprenticeships, Skills, Children and Learning Act 2009: 31 March 2020

9 Consultation decisions: Exceptional arrangements for exam grading and assessment in 2020

together a panel of assessment and statistical experts to advise on technical issues in addition to working with technical colleagues from the exam boards.

The great majority of students who had been entered to take exams this summer have received a grade calculated by the exam board for each of their subjects. Students who feel that the grade does not reflect their ability or those for whom it was not possible to issue a calculated grade, will be able to take exams in the additional exams series this autumn or, if they prefer, next summer.

This report details the technical considerations and decisions that were necessary to award grades to students this summer in the absence of the opportunity for formal assessment. An analysis of the results students received following exam boards' delivery of that process for AS and A level qualifications is also presented. Similar analyses covering GCSE qualifications will be published on 20 August in line with GCSE results day.

2.2 Scope

The Secretary of State's statements and his direction to us[8] covered GCSEs, AS and A levels. We regulate Extended Project Qualifications and Advanced Extension Awards in a similar way to that in which we regulate those qualifications. Extended Project Qualifications (EPQ) and the Advanced Extension Award (AEA) in maths are also used for entry to university or employment. We therefore decided that the exceptional arrangements we were to put in place for GCSEs, AS and A levels would also apply to the EPQ and the AEA in maths.

We put in place different arrangements for other qualifications we regulate, including other general qualifications, such as the Cambridge Pre-U and the International Baccalaureate Diploma and for vocational and technical qualifications. These are not covered by this report. We do not regulate International GCSEs and we did not oversee the arrangements put in place for them this summer.

Important context for some of the technical discussion that follows is the recent programme of qualification reforms. GCSE, AS and A levels have recently been reformed in England. The introduction of new specifications was phased, with subjects being introduced over a number of years. The first awards of new AS qualifications took place in 2016 and the first new GCSE and A levels in 2017. The first award of the final few reformed GCSEs is to take place this summer. In 2017 we published a timeline that set out subjects that were in each phase of the reform programme.[10] The potential impact on the phasing of different subjects in relation to reform was considered when putting in place the arrangements described in this report. Where the phasing of reform is important, this is drawn out at the relevant point.

As the final phase of general qualifications reform was due to complete this year, summer 2020 would have been the first summer in which all qualifications in England followed a linear structure. This meant that students had not taken any formal assessments prior the summer. In some subjects, students will have completed, or partially completed, non-exam assessment (NEA) tasks, however,

[10] https://www.gov.uk/government/publications/timeline-of-changes-to-gcses-as-and-a-levels/changes-to-gcses-as-and-a-levels-that-will-affect-each-current-school-year-group

where these tasks had been completed, they will not have been subject to the relevant external moderation processes delivered by exam boards.

The qualifications and subjects to which the arrangements discussed in this technical report relate are listed in Annex A. They comprise 53 full-course GCSE, 2 short-course GCSEs, 43 AS and 57 A level subjects, as well as the Extended Project Qualification and AEA maths. The numbers of students and entries for each qualification are shown in Table 2.1.

Table 2.1 Number of students and entries for relevant qualification this summer[11]

	Summer 2020		
	Number of students	Number of entries	Awarding organisations
GCSE	951,279	4,771,415	AQA, OCR, Pearson, WJEC
AS	37,262	70,505	
A level	274,567	718,276	
EPQ	35,611	35,611	AQA, ASDAN, City and Guild, OCR, Pearson, WJEC
AEA in maths	215	215	Pearson

2.3 Centre assessment grades and rank orders

As exams did not take place this summer, students' grades were instead based on evidence of their likely performance in the exams had they gone ahead. In the absence of any formal assessments delivered by exam boards, evidence relating to students' expected performance in qualifications this year was required.

In April 2020 we published guidance for Heads of centres, which was updated and republished on 22 May.[12] Schools and colleges (centres[13]) were asked to submit to exam boards, for each student and for each subject for which they were entered, the grade they judged the student would most likely have received had the exams taken place (the centre assessment grade or CAG), and the rank order of each student at each grade in each subject in that centre. While the approach to awarding grades this summer was still at the early stages of consideration, we judged that both sources of evidence would be needed to determine final grades. Rank order

[11] Note that the A level, AS, EPQ and AEA mathematics figures are based on data submitted to Ofqual by exam boards around one week before results were issued so may differ slightly from the JCQ published figures. The GCSE data are more provisional and based on data submitted during the live awarding process.

[12] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/887018/Summer_2020_Awarding_GCSEs_A_levels_-_Info_for_Heads_of_Centre_22MAY2020.pdf

[13] The term centre is used throughout this report to refer to any organisation undertaking the delivery of an assessment to students on behalf of an awarding organisation. In the context of general qualifications, these are typically schools and colleges but may include other types of institution.

information would provide a more fine-grained ‘scale’ than using grades alone. Feedback from teacher representatives also suggested that this made the task more intuitive for teachers, who were used to grading their students and could more easily rank within a grade than overall, particularly in larger centres.

The guidance set out the various sources of evidence that centres should consider. It also emphasised the importance that judgements were based on objective evidence, to minimise the risk of any unconscious bias influencing those judgements. This is of particular relevance in relation to the potential impact on protected groups and so, following engagement with equalities organisations, we provided much more guidance to Heads of centre in the updated version of the guidance published in May. The guidance also made clear that schools and colleges should not discuss their evaluation of the evidence or disclose the judgements they reached with students or their parents or carers. This was to enable teachers to make their judgements fairly and without being put under unreasonable pressure.

Judgements made by individual teachers were to be signed off by at least two members of teaching staff, one of which was to be the head of department. In addition, heads of centre were required to submit a declaration that the grades and rank orders being submitted were correct and had been generated according to the guidance.

Given that schools had been closed from 18 March and England was in lockdown, we had no opportunity to standardise teachers before they made their judgements. In line with the direction from the Secretary of State, we were clear that the CAGs would be standardised using a statistical model that would take account of the historical results in that subject in the centre, and the prior attainment profile of the cohort of students taking that subject compared to previous years.

How exam boards used the evidence collected from schools to award grades this summer is described in detail in Section 8.

3 The case for standardisation

As outlined above, the most readily available form of evidence regarding students' recent performance is teacher estimates. However, evidence suggests that teacher estimates – despite the best intentions – are not without limitations. Such limitations are particularly important to recognise when the scale of the GCSE, AS and A level system – and the number of individuals involved – is considered. For example, in 2019, over 700,000 A level results and over 5 million GCSE results were issued in England.

It is also important to consider that, in the context of current events, the period of time between exams being cancelled and teachers needing to generate and supply estimates to exam boards (to allow results to be issued in August) was very short. As such, although guidance was provided to centres to assist them in making their judgements (including how to avoid any unconscious bias), it was not possible to standardise the process between different schools and colleges. This would be challenging in normal circumstances but is particularly challenging given current events. For example, public health restrictions meant that schools and colleges were closed for normal teaching, and some teachers will have been directly affected by on-going events. This, alongside evidence relating to the accuracy of teacher estimates, points to the need for a process of standardisation to be used.

3.1 Forms of accuracy in teacher estimates

There are two key aspects to accuracy, and it is important to distinguish between them when considering the ability of teachers to estimate students' grades:

1) **Absolute accuracy** – relates to the ability of a teacher to estimate the actual grades that individual students will achieve
2) **Relative accuracy** – relates to the ability of a teacher to estimate the rank order of students by their actual grades, i.e. their levels of achievement relative to one another

Using an extreme example (see Table 2.2) to illustrate the point, a teacher might be overly optimistic about the grades that all of their students will achieve and get none of the individual estimates correct (absolute accuracy = 0.0). Nonetheless, they may, *at the same time*, judge perfectly the rank order of their students' achieved grades (relative accuracy = 1.0).

Table 2.2. Estimated and actual grades for students where the teacher has estimated the rank order of students perfectly (relative accuracy) but has overestimated every student's actual grade (absolute accuracy).

Student	Estimated grade	Actual grade
1	A*	A
2	A	B
3	B	C
4	C	D
5	D	E
6	E	U

The distinction is important because relative accuracy allows for standardisation/moderation of teacher estimates in the absence of absolute accuracy. Without this information, only coarse adjustments, changing grades for all students on a particular grade, would be possible, leading to the over/under adjustment of grades. With it, it is possible to make grading adjustments that are sensitive to differences between, for example, a borderline, a middling, and a strong grade C student.

Ordinal judgement as a basis of psychological measurement (including educational assessment) has its roots in psychophysics[14] and has become an established method for conducting comparability studies (comparing students' performances across different examinations)[15] and rating students' performances as an alternative to traditional marking.[16] Laming[17] has argued that all judgements are ordinal – a comparison of one thing with another – and that people are typically more accurate at making comparisons between two *concrete* things than between a *concrete* thing and an *abstract* thing. This is not least because abstractions will differ between people according to their experiences. In the case of estimated grades, it is easier – and consequently more reliable – to compare the performances of students with one another than it is to compare the performance of students with the abstract notion of one or more grade performances.

[14] Thurstone, L.L. (1927b). A law of comparative judgment. Psychological Review, 34, 273–286. Chapter 3 in L.L. Thurstone (1959), The measurement of values. Chicago, Illinois: University of Chicago Press.

[15] Bramley, T. (2007) In: P. Newton, J. Baird, H. Goldstein, H. Patrick, and P. Tymms (Eds.), Techniques for monitoring the comparability of examination standards, 246-294. London: QCA

[16] Pollitt, A. (2004, June). Let's stop marking exams. Paper presented at the annual conference of the International Association for Educational Assessment, Philadelphia, USA.

[17] Laming, D. (2004). Human judgment: The eye of the beholder. London: Thomson.

Recent and historical research[18] examining the estimated grades provided by teachers[19] have found consistent tendencies in terms of the levels of accuracy, and further support both: (i) the need to standardise/moderate estimated grades; and (ii) doing so on the basis of ordinal estimates.

Key observations from these studies are:

- Teachers tend to be optimistic in their expectations approximating the following pattern:
    - Optimistic $\cong {}^{1}/_{3}$ of estimates
    - Accurate $\cong {}^{1}/_{2}$ of estimates
    - Pessimistic $\cong {}^{1}/_{6}$ of estimates
- The correlation between teacher estimates and actual grades is relatively strong with correlations between 0.76 and 0.85[18]. This means that teachers can estimate the *relative* performance of their students within their class with high accuracy, even if the *absolute* accuracy is lower.
- Overall, these correlations show a tendency to be more accurate recently than they have been historically, with weaker correlations (between 0.45 and 0.79) common in studies up to 1997. While not documented in the research literature, this improvement over time may be due to the improved diagnostic information available to teachers – including the unitisation of A levels from 2000 (and later GCSEs) offering feedback on in-course progress – combined with their increased personal accountability for results data.
- Despite this overall trend of improved relative accuracy, there is some evidence suggesting that relative accuracy has reduced at A level since the introduction of the (linear) reformed qualifications. While yet to be proven, it is intuitive to relate this to the decoupling of AS from A level, since AS no longer contributes to the A level outcome and the reduced uptake means it is a less prevalent source of evidence for prediction.
- The absolute and relative accuracy of estimates vary by subject but not on the basis of observable features. In the current context, this may also be affected by the differential phasing of qualification reform, assuming that predictions are less reliable in the first year of a qualification.

[18] Dhillon, D (2005) Teachers' estimates of students' grades: Curriculum 2000 Advances Level Qualifications. British Educational Research Journal 31(1) 69-88.

Methods used by teachers to predict final A Level grades for their students
The accuracy of forecast grades for OCR GCSEs in June 2014
The accuracy of forecast grades for OCR A levels in June 2012
Improving the Higher Education Applications Process

[19] It is important to note that the majority of work in this area has focussed on 'predicted' grades at A level, due to their higher stakes arising from use in university admission processes. The use of these predicted grades as opposed to estimates potentially used for internal purposes in centres may have impacted on these findings. However, given the current context, it is not unrealistic to expect similar tendencies to be present in teacher estimated grades.

It is useful to consider the potential impact of over-estimates on outcomes. At GCSE this is perhaps best evidenced by the work performed during the investigation of GCSE English grading in 2012. To provide context, an historical analysis of the relationship between estimated and actual GCSE English grades was performed for earlier years unaffected by the subject of the investigation. The analysis, based on 332,300 students entering with AQA who had estimated grades, showed that, were students to have been awarded their estimates in 2011 rather than those that they actually achieved, the percentage of students achieving grades A, C and F (and above) would have been as follows:

	Grade A	Grade C	Grade F
Estimated grades	20.92%	77.24%	99.06%
Actual grades	17.56%	65.23%	97.53%
Cumulative Percentage Difference	+3.36%	+12.01%	+1.53%

This equates to approximately 40,000 more students being awarded a grade C based on their estimated grade compared to those they actually achieved. This figure is likely high due to the stakes around achieving a grade C (including for the school due to accountability measures at the time), combined with a peak in the numbers of students close to this grade.

It is also important to note that the evidence presented above was based on estimated grades generated as part of business as usual from a year unaffected by the issue that was being investigated. And, while summer 2020 results are not being used for the purposes of school accountability[20], evidence suggests that estimated grades will tend towards over-estimation. This is merely human nature but is particularly likely to be the case given the prominent role that these estimates have in the awarding of grades this summer.

3.2 Equalities considerations

It is also important to consider the implications of using teacher estimates from an equalities perspective. In April, we published a review of the literature[21] considering the nature and extent of any bias that might arise in CAGs this summer.[22] In summary, studies of potential bias in teacher assessment suggest that differences between teacher assessment and exam assessment results can sometimes be linked to student characteristics, including gender, age within year group, ethnicity, special educational needs, and having English as an additional language. However, such effects are not always seen, and when they are, they tend to be small and inconsistent across subjects.

The accuracy of teachers' estimates has been considered when examining the accuracy of teachers’ A level grade predictions for students' university admission

20 https://www.gov.uk/government/publications/coronavirus-covid-19-school-and-college-performance-measures/coronavirus-covid-19-school-and-college-accountability

21 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879605/Equality_impact_assessment_literature_review_15_April_2020.pdf

22 https://www.gov.uk/government/consultations/exceptional-arrangements-for-exam-grading-and-assessment-in-2020

applications, and in other research using individual exam board data to examine the accuracy of the GCSE and A level grades that boards previously collected from teachers. The same distribution of exactly accurate and over-/under-predictions, and pattern of attainment-dependent prediction accuracy, have been found in both strands of research. Findings on individual variables are also broadly similar: subject has a small but unsystematic effect; gender and age have small effects that are inconsistent across subjects; centre type has a small effect that can be attributed to the ability of the students attending different types of centres.

The literature also suggests there are likely some effects on prediction accuracy of ethnicity (that is, more over-prediction for some ethnic minority groups) and disadvantage (that is, more over-prediction for the more disadvantaged students in general, and less over-prediction for the more disadvantaged students among high attainers). Further work is, however, necessary to properly estimate these effects.

To support teachers in making objective judgements we provided additional guidance on how to avoid unconscious bias in decision making[23].

3.3 Summary

In summary, without standardisation, the likely inflationary effects of using unstandardised teacher estimates would undermine confidence in the grades awarded to students this summer. And, while evidence suggests problematic levels of absolute accuracy in teacher estimates, it is likely that there will be reasonably high levels of relative accuracy. This suggests that overcoming the tendency for the estimates to be optimistic would lead to a legitimate basis for grading students.

It is also important to consider that, without standardisation, any differences between the standard applied by centres would not be addressed. It is likely that some centres will have been more optimistic than others when submitting their CAGs, and indeed some centres may have been pessimistic. Applying a method of standardisation aims to ensure that, as far as possible, such differences are addressed, increasing the fairness to students.

[23] Guidance for Heads of Centre, Heads of Department and teachers on objectivity in grading and ranking

4 Principles of standardisation

Given the importance of standardisation it is helpful to establish underpinning principles to support the development of the approach. To ensure these principles reflected the widest available range of views from stakeholders, they formed part of a policy consultation which ran from 22 May 2020 to 8 June 2020. This section outlines these principles.

4.1 Aims of standardisation

To design and evaluate potential approaches to standardisation it is important to define a set of aims. A proposal for these aims was subject to consultation and was broadly supported by respondents (89% of respondents either agreed or strongly agreed). A slight refinement was made post-consultation to reflect the comments received. This was to make clear that the transparency and ease of explanation of the approach was a lower priority than other aims.

The confirmed aims of the standardisation process are therefore:

i. to provide students with the grades that they would most likely have achieved had they been able to complete their assessments in summer 2020
ii. to apply a common standardisation approach, within and across subjects, for as many students as possible
iii. to protect, so far as is possible, all students from being systematically advantaged or disadvantaged, notwithstanding their socio-economic background or whether they have a protected characteristic
iv. to be deliverable by exam boards in a consistent and timely way that they can quality assure and can be overseen effectively by Ofqual
v. to use a method that is transparent and easy to explain, wherever possible, to encourage engagement and build confidence

It was anticipated that during the development of the approach there would be times when these aims would be in tension. For example, where accuracy might need to be balanced against taking a common approach for as many students as possible. In these instances, an optimal balance was sought.

The first of the aims – to deliver the grades which students would otherwise have achieved – is fundamental but challenging. While a degree of assessment unreliability always exists in any qualification system, and leads to students being awarded grades other than their ‘true grade’[24], it is likely that in the absence of exams and assessments, it will be more challenging to award reliable grades this summer.

It is inevitable that, whatever method is applied, some students will be awarded a grade that is lower or higher than that which they would otherwise have achieved had exams gone ahead. Given the overarching purpose of awarding grades this summer is to provide results to allow student progression, the under-awarding of

[24] https://www.gov.uk/government/publications/reliability-of-assessment-compendium

grades to a significant proportion of students would have far more significant consequences.

Given the heightened risk of students being under-graded, in a context where they have not had the opportunity to complete their assessments, the risk of unfairness is significant. This might suggest it is appropriate to award outcomes this summer that are higher than might be expected in order to protect the interests of the current cohort, while accepting that many other students will be consequently over-rewarded.

There are, however, drivers in the opposite direction. In his direction to Ofqual on 31 March 2020, the Secretary of State asked that "Ofqual should ensure, as far as is possible, that qualification standards are maintained and the distribution of grades follows a similar profile to that in previous years"[25]. Were generosity to be built into the system this year to protect those who would otherwise be disadvantaged, an inflation in outcomes would be inevitable. Further, for students who positively benefit from both unreliability of the process and a lenient overall award, the effect might be particularly significant.

This presents risks to:

- the credibility of the grades issued this summer
- the inferences that can be drawn by users of the qualification outcome
- fairness to cohorts of students over-time, potentially disadvantaging those who certificated last year and those who will certificate next year if in competition with the 2020 cohort for admissions or employment
- the healthy operation of other areas of the system, such as HE, that are usually reliant on stable overall qualification outcomes

It was decided, where possible, to seek to maintain overall qualification standards and to produce overall outcomes broadly in line with those of previous years. However, during the design of the standardisation process and its operation, there were several decision points which presented the opportunity to give benefit of the doubt to students. It was decided, notwithstanding the desire to broadly maintain standards, that these opportunities for leniency would be taken where possible. These are outlined where appropriate within this report.

It was also recognised that there would be occasions where producing a distribution of grades similar to previous years would be challenging. Specifically, where the size of the cohort entered into a qualification either nationally or at centre level was so low as to make statistical approaches to the maintenance of standards inappropriate.

25 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877611/Letter_from_Secretary_of_State_for_Education_to_Sally_Collier.pdf

4.2 Balance of evidence

Where available, the process of standardisation can draw on the following sources of statistical evidence:

- historical outcomes for each centre
- the prior attainment (Key Stage 2 or GCSE) of this year's students and those in previous years within each centre
- the expected national grade distribution for the subject given the prior attainment of the national entry[26]

Approaches to standardisation vary in the emphasis they place on this statistical evidence versus the submitted CAGs. We consulted on the relative weight that should be given to these two sources of evidence.

An approach that places more weight on CAGs would assume that these are correct unless there is statistical evidence to the contrary. This is intuitively appealing because it would minimise the number and scale of any changes between CAGs submitted and final calculated grades issued. It would also more strongly reflect the professional judgement of centres as to the grade worthiness of their students. However, we judged that it would have significant downsides.

Firstly, it is likely that differences in the standards applied by different centres would persist, which would be unfair. In particular, students from centres giving severe CAGs would be disadvantaged compared to students from centres giving accurate CAGs who would, in turn, be disadvantaged compared to students from centres giving generous CAGs. Indeed, there is evidence that different centres understandably took different approaches to the creation of CAGs and consequently that different standards were applied across centres.

Secondly, it is likely that this would produce results that were overall very significantly increased, standards would not be maintained and the currency of the qualifications would be called into question. Research shows when teachers predict / estimate grades for other purposes such as for university entrance or to inform the grade boundary setting process, they tend to be too generous more often than they are too severe[27]. This would be unfair to students in the previous and forthcoming years with whom this year's students will compete for opportunities in education, training and employment. This would also undermine the credibility of the grades awarded. Indeed, the CAGs were found to be consistently and significantly generous compared to previous years outcomes (see Section 9).

We proposed an approach that placed more weight on statistical expectations - determining the most likely distribution of grades for each centre based on the previous performance of the centre and the prior attainment profile of this year's students. Then using the submitted rank order to assign grades to individual students in line with this expected grade distribution. We noted that the likely result of this approach would be that the final calculated grades received by centres would more often differ from those submitted. Indeed, while most CAGs were unchanged, a significant proportion were (see Section 9).

26 Drawing on results of the NRT if/as relevant
27 Exceptional arrangements for exam grading and assessment in 2020

We judged that there were a number of advantages to this approach. Firstly, such an approach would reflect research evidence about the likely accuracy of CAGs versus rank orders.

Secondly, this approach would be more likely to ensure that a consistent standard is applied across centres, being more effective in removing generosity/severity, and so is more likely to be fair to students.

Thirdly, this approach makes it more likely that the resultant national grade distribution would match that expected given the prior attainment of this year's cohort. Moreover, this could be achieved without unfairly disadvantaging students from those centres providing severe or accurate CAGs compared to students from centres providing generous CAGs.

Therefore, we proposed that an approach that placed more weight on statistical expectations is appropriate and most fair to students, particularly in light of Ofqual's statutory objective to maintain standards over time. 54% of consultation responses agreed or strongly agreed. We recognised though that in certain circumstances (such as for centres with low entries and low entry subjects), it would be appropriate to place more weight on CAGs than previous centre performance. We therefore decided therefore to adopt a modified form of our original proposal, whereby, the statistical standardisation model should place more weight on historical evidence of centre performance (given the prior attainment of students) than the submitted CAGs where that will increase the likelihood of students getting the grades that they would most likely have achieved had they been able to complete their assessments.

4.3 The role of trajectory

Approaches to standardisation can also vary in whether they carry forward the outcomes that centres have achieved in previous years (taking changes in the prior attainment of the students into account) or seek to reflect any trends of improvement or deterioration in outcomes over previous years (the trajectory of the centre). In the latter approach, the statistical model would predict higher grades for those centres on an upward trajectory and lower grades for those on a downwards trajectory. In other words, an assumption would be made that the trend in results which has occurred over recent years will continue.

If we could have been confident in predicting continued trends in centres' results for 2020 the clear advantage of a centre trajectory approach is that it would lead to fairer awarding. Analysis performed in 2018[28] considered the variability of GCSE outcomes for schools and colleges. This showed that the results for the vast majority of centres do not show a consistent trajectory year-on-year. In 2015 and 2016, 90% of centres were classed as having 'stable' outcomes and 8.5% of centres were classed as having 'unstable results' that either improved and then deteriorated or vice versa. Only 0.8% of centres had results that increased by more than the national average change in both 2015 and 2016 and only 0.5% of centres had results that decreased more than the national average change in both 2015 and 2016. This lack of stability over time in improvements or deteriorations in performance for the overwhelming

28 What causes variability in centre level GCSE results year-on-year? Some further analysis

majority of centres means that any statistical model may be unreliable in predicting trends in performance in 2020.

Given this unreliability it might seem fairest to only apply the trajectory model to those centres who are predicted to improve and not to those predicted to decline. Although intuitively appealing, this would advantage students in these centres compared to those in all other centres (on the basis of a predicted trajectory that might in reality not have come to pass). It could also lead to significant inflation in national outcomes which may undermine the credibility of the grades awarded to students this summer. Any allowance made would need not to conflict with Ofqual's statutory objective to maintain standards over time.

We consulted on whether standardisation ought to seek to reflect the trajectory of centres. We proposed that the model should not seek to do so. 45% of respondents strongly agreed or agreed with our proposal (school or college respondents were more likely to agree than teachers responding in a personal capacity or parents/carers) and a sizeable percentage (29%) of respondents neither agreed nor disagreed. We decided therefore that the standardisation process would not seek to reflect centre trajectory in predicting results.

4.4 Correcting for potential bias in centre assessment grades

The potential for inequalities caused by bias in the system of assessment has been at the forefront of our thinking throughout the design and operation of this year's process. Students must be rewarded based on their ability relative to the construct being assessed.

Centres were asked to use their professional experience to make a fair and objective judgement of the grade they believed a student would have achieved had they sat their exams and completed any non-exam assessment. However, research evidence shows that bias can occur in similar situations such as in predicting grades for university entrance and estimating grades for use in standard setting; although the findings are mixed[29] - the effects appear to be variable and context dependent. So, to support teachers in making judgements we provided additional guidance on how to avoid unconscious bias in decision making[30].

Assuming that centre level demographics are stable over time, the standardisation approach should not exacerbate attainment gaps between students with different protected characteristics or from different socio-economic backgrounds. Outcomes for centres would be largely maintained from previous years (where the ability of their cohort has not changed). Hence, the relationship between results and any centre-level demographics would also be largely maintained from previous years when assessment occurred in the normal manner – by exam and / or non-exam assessment.

29 Equality impact assessment: literature review
30 Guidance for Heads of Centre, Heads of Department and teachers on objectivity in grading and ranking

Further, in choosing the statistical standardisation model the relative impact on centres with different characteristics (such as socio-economic status (SES), special educational needs and/or disabilities (SEND), ethnicity and gender) was carefully considered and evaluated by comparing differences in outcomes to those observed in previous years (see Section 7.5).

However, given that it was possible that some degree of bias might exist in CAGs, the potential to statistically adjust the standard being applied to different groups of students (for example by SES, SEND, ethnicity and gender) was consulted on. This would be done to reproduce historical patterns of results for the different groups – replicating attainment gaps. We proposed not to use this approach to correct for any bias and most respondents (64%) strongly agreed or agreed, and 17% disagreed or strongly disagreed.

We proposed not to use the approach because it would be challenging to identify whether individual centres did or did not submit CAGs and rankings which were affected by bias, and, even were it possible to do so, it would not be possible to identify the extent or impact of any such bias.

Further, the approach would have the effect of changing the rank order of students within each centre – promoting/demoting some students over others on the basis of their SES, SEND, ethnicity or gender, for example. The final calculated grades received would no longer reflect the centres' judgements about the relative grade worthiness of individual students.

We committed to not changing the rank orders submitted by centres because we believe centres are best placed to judge the likely performance of their students relative to one another at the end of the course.

Further, such an intervention would be unprecedented in the history of qualifications in England. For example, when qualifications are reformed, the impact of changes to assessment models on those students with protected characteristics are evaluated and carefully considered. However, there is no attempt during the standard setting process to maintain previous differences in outcomes between students who share particular protected characteristics, such as disability, ethnicity or gender, or who are from particular socio-economic backgrounds. This is in part due to the difficulty of establishing the extent to which differences in outcomes are caused by differences in educational opportunity rather than changes in the assessment models.

These considerations were important, however, evidence from our analysis of calculated grades shows no evidence of systemic bias in the grades awarded this year (see Section 10).

5 The maintenance of standards

5.1 The importance of maintaining standards

The maintenance of standards is fundamental to the role of Ofqual, as articulated in our statutory objectives.[31] It is crucial for ensuring fairness to students – both in terms of students taking qualifications with different exam boards in the same year, and students taking the same qualifications over time. This year is no different to any other in this regard; our approach aims to ensure that, as far as possible, there is a level playing field for students. In particular, we aim to ensure that it is no easier to get a particular grade in a subject with one exam board than with another, or in a different examination series.

The importance of maintaining standards is clear when the uses to which results are put are considered. Within society, grades are a form of currency, and their use relies on there being stability over time (unless there is a clear reason why results might change). If standards are not maintained, then the value and credibility of grades is likely to be undermined. This would be problematic given that there is a reliance within other parts of the education system (and more widely) on qualification grades being comparable over time.

Students' grades are used for a variety of purposes, including giving access to further education, training and employment opportunities. This is particularly the case for A levels, since the grades that students achieve are used by many students to access higher education (HE) opportunities at Universities and Colleges. If standards are not maintained, this can create challenges for HE admissions – both in terms of differentiating between students of different abilities, and in ensuring that students are admitted to courses that are suitable for them. Similarly, GCSE grades are often used for allowing access to further (A level) study, training or employment.

Maintaining standards is important in any year, and it is no less the case this year. Yet, current circumstances mean that the challenges of maintaining standards this year are unique. A vital piece of evidence used for the maintenance of standards is statistical evidence indicating the expected attainment of students at the cohort level in each subject. The following section sets out how statistical evidence is used to support the maintenance of standards in a typical year, and how that evidence is being used this summer. While the use of statistical predictions at the cohort level is broadly similar to other years, there are some differences, as outlined below.

5.2 The role of statistical predictions

For general qualifications (eg GCSE, AS and A levels), standards are maintained in a typical year using a combination of statistical and judgemental evidence.[32] The over-arching principle is that if the cohort of students sitting a given qualification in one year is similar to the cohort in another year, then the outcomes should also be

31 Apprenticeships, Skills, Children and Learning Act 2009

32 Taylor, R. and Opposs, D. (2018) 'Standard setting in England: A levels'. In Baird, J., Isaacs, T., Opposs, D. and Gray, L. (eds) (2018) *Examination standards: how measures & meanings differ around the world*. London: UCL IOE Press

similar. In practice, this is operationalised by using prior attainment-based predictions and judgemental evidence from senior examiners.

Prior attainment-based predictions provide a mechanism for facilitating alignment between exam boards and can be used to evaluate the extent to which standards are aligned. The predictions model the relationship between prior attainment and outcomes in a previous examination series – the reference year(s)[33] – then use this relationship to predict outcomes for the current cohort of students (given their prior attainment). For AS, A level and other level 3 general qualifications (eg the EPQ), the predictions use mean GCSE grade as a measure of prior attainment. For GCSEs, the predictions use mean KS2 score as a measure of prior attainment. Given that the predictions model the relationship between prior attainment and outcomes, they only include those students that are matched to their prior attainment (known as 'matched' students). Predictions are more reliable the greater number of matched students there are, so in a typical year they are only used to evaluate exam boards' awards where there are more than 500 matched students for a given specification.

The statistical predictions are generated at the cohort level for a specific group of students, typically those that would be expected to certificate in that qualification – 18-year-olds for A level and the EPQ, 17-year-olds for AS, and 16-year-olds for GCSE. The predictions include all students (of the appropriate age) that are matched to their prior attainment, except for GCSE where students from independent and selective schools are excluded. This is because evidence suggests that students from independent and selective schools perform differently to the overall cohort (ie they have a different relationship between prior attainment and outcomes). As such, if the proportion of students from these centres changed over time or within a particular exam board, then the predictions may under- or over-predict outcomes.[34] We do not think that this would be fair, hence these students are routinely excluded from the predictions.

There are 2 main steps to generating predictions.[35] First, an 'outcome matrix' is generated for the reference year(s) for each subject and qualification (see Figure 5.1 for a fictitious A level outcome matrix). Students in the reference year(s) that are matched to their prior attainment are divided into deciles based on their prior attainment at GCSE, and a matrix is created that shows how the students in each decile went on to perform in that subject. The top decile (numbered 1 in Figure 5.1) includes the most able students, and the bottom decile includes the least able students. The percentages are cumulative such that, in this example, 80% of students in the top decile achieved a grade B or above in this subject in the reference year(s). Note that there is a separate outcome matrix for each subject and at GCSE students are divided into octiles according to their prior attainment, rather than deciles.

[33] The reference year can be a single year or multiple years. Typically, the reference series is the first 2 years that a new specification was awarded.

[34] Predicting GCSE outcomes based on candidates' prior achieved key stage 2 results

[35] See also our blog, prediction matrices explained

		A level grade						
		A*	A	B	C	D	E	U
Mean GCSE decile	1	70	75	80	85	90	95	100
	2	65	70	75	80	85	90	100
	3	60	65	70	75	80	85	100
	4	55	60	65	70	75	80	100
	5	50	55	60	65	70	75	100
	6	45	50	55	60	65	70	100
	7	40	45	50	55	60	65	100
	8	35	40	45	50	55	60	100
	9	30	35	40	45	50	55	100
	10	25	30	35	40	45	50	100

Figure 5.1. Example outcome matrix

Using this outcome matrix, it is then possible to predict how students in the current year are expected to perform, given their own prior attainment. For example, using the outcome matrix above, 70% of students in decile 1 would be expected to achieve a grade A*; 75% would be expected to achieve a grade A or above; 80% would be expected to achieve a grade B and above; and so on. This is repeated for each decile and then aggregated together to form a prediction that shows the probability of the current cohort achieving each grade, given their prior attainment. This means that if the prior attainment profile of the cohort is stronger in the current year compared to the reference year(s), then outcomes would be expected to be higher.

Outcomes matrices are generated at the national level, using the results of all matched students taking qualifications in the relevant subject. This provides a common statistical model on which all qualifications in that subject can be based. In a typical year, exam boards generate a prediction for each of their own specifications, based on their own cohort of students who are matched to their prior attainment, using that same (national) relationship. This approach supports a common alignment of standards across exam-boards in a given year and supports the maintenance of standards over time. Because the predictions reflect the prior attainment profile of the students taking a given specification in any given year, one exam board might have a higher prediction than another if the prior attainment profile of their students is stronger. The predictions are used by exam boards, alongside senior examiners' judgements of the quality of students' work, to set grade boundaries for each specification.

Although predictions are usually generated for all grades, in a typical year they are only used to inform the setting of grade boundaries – alongside examiner judgement – at key grades. The key grades are A*, A and E for A level; A* and E for the EPQ; A and E for AS; and 9, 7, 4 and 1 for GCSE. The remaining grade boundaries are then calculated arithmetically – for example, the grade 8 boundary is set halfway between the grade 9 and 7 grade boundaries. This enables sufficient statistical oversight of outcomes and the opportunity to thoroughly scrutinise the quality of students' work in a way that would not be possible or appropriate at every grade boundary.

5.2.1 Using statistical predictions during reform

Maintaining standards when qualifications have been in place for a sustained period of time can be a technically complex task. However, where there is a change in the design of qualifications, the challenge of setting and maintaining appropriate standards increases. As noted in Section 2.2, over recent years, general qualifications in England have undergone an extensive programme of change, and it is important that there is sufficient confidence in the standards set during this period of reform.

Predictions are an important source of evidence for maintaining standards and ensuring alignment between exam boards in any year, but they are particularly important during periods of reform. We know that when qualifications change there is generally a small dip in performance when the qualification is first awarded, and this can make it more difficult for senior examiners to judge the quality of student work. This is known as the saw-tooth effect,[36] and is due to teachers being less familiar with the new qualifications and there being fewer past papers and other support materials available. Performance then tends to gradually improve over subsequent assessment sittings.

During the first and second awards of the recently reformed GCSE, AS and A level qualifications, we required exam boards to prioritise the use of statistical evidence to ensure that standards were maintained. This ensured that students were not disadvantaged due to being in the first cohorts to sit the reformed qualifications. To do otherwise would have risked disadvantaging students simply due to the academic year that they happened to be in.

5.2.2 Using predictions to evaluate awards

In a typical year – and where entries are sufficiently large[37] – exam boards report data to us from each of their awards showing the outcomes for matched students compared to the predictions at each of the key grades. We then evaluate the outcome of each award against a given reporting tolerance (that depends on entry size). Where the actual and predicted outcomes for a specification differ beyond this reporting tolerance, the exam board must inform us and provide additional information to support an out of tolerance award. We use tolerances in a typical year because it is rarely possible to meet the statistical prediction exactly (this depends on the mark distribution and the number of students on each mark) and reflects the role of examiner judgement in awarding.

In the first and second years of awarding the recently reformed GCSE, AS and A level qualifications, we did not use reporting tolerances and instead required exam boards to meet the predictions as closely as possible (unless there was compelling evidence not to do so). As described above, this was to avoid students being disadvantaged due to being in the first cohorts to sit the reformed specifications. From the third year onwards, there is a greater role for examiner judgement and the tolerances are used to evaluate exam boards' awards.

[36] Investigation into the Sawtooth Effect in GCSEs, AS and A levels

[37] Where the entry numbers are relatively small, exam boards will balance the use of statistics with the judgements of their senior examiners.

5.3 Arrangements for summer 2020

Given that there were no assessments this summer, the challenges of maintaining standards are unique, since no formal evidence of student performance is available. Nonetheless, the role of statistical predictions remains key. In the same way that predictions are used during periods of reform to ensure that any dips in performance are compensated for, predictions are an important source of evidence this summer when there is no evidence of student performance. The use of predictions aims to ensure that students are not advantaged or disadvantaged due to current circumstances – or relative to those sitting qualifications in other years.

It is worth noting here that the precision to which the predictions can be met differs to a typical year. This is because of the way that the standardisation model operates, so is considered further in Section 8. It is also important to draw a distinction between the statistical predictions that are used at the cohort level (that are used every year), and the evidence that is being used to standardise the outcomes for centres this summer. As outlined above, the statistical predictions are used at the cohort-level to support the overall maintenance of standards in a qualification. And, while there are some similarities with how prior attainment is used as part of the standardisation process, the two aspects of setting standards should be considered as distinct. This section considers the use of statistical predictions to maintain overall outcomes in qualifications this year. The details of the centre-level statistical analysis relating to standardisation will be given in Section 8.

Broadly, the qualification-level statistical predictions this summer are generated using the same approach as in any year, as described above. And, as in previous years, we have consulted with exam boards about the details of those predictions – for example, the reference year(s) that should be used and the cohort that predictions should be generated for. We published the approach for generating predictions (as in any year) ahead of exam boards starting the awarding process. This year those details are included in Annex D of our 'Requirements for the calculation of results in summer 2020'.[38] Given current circumstances, there are some differences though to a typical year, and the following sub-sections draw out the key points relating to the use of statistical predictions this summer.

5.3.1 National approach to standard setting and evaluating awards

Most general qualifications are offered by multiple exam boards. Centres choose which specification they wish their students to study and therefore the exam board with which they enter. While the content of the specifications is broadly the same, there are generally differences in the structure or style of the assessments. For example, some GCSE maths qualifications have 2 components, while others have 3 components.

In a typical year, exam boards award their own specifications independently. As such, each exam board generates their own statistical predictions for their cohorts of students, using the national relationship between prior attainment and outcomes from the reference year(s). This information, alongside senior examiner judgements of students' work, is used to set the grade boundaries. We then monitor the

38 Requirements for the calculation of results in summer 2020

outcomes of each exam board's awards to ensure that it is no easier to get a grade with one exam board than another.

The importance of ensuring comparability between exam boards is no different this summer. Typically, differentiation between exam boards is based on the performance of the students on the assessments, which may differ depending on the demand of the assessments. For example, grade boundaries may differ between exam boards to ensure that overall standards are comparable. This evidence is not available this summer, but it is important that students are treated equitably regardless of the exam board that they are entered with.

As different assessments are not available across exam boards this summer, rather than requiring exam boards to generate predictions separately for their own cohorts we have required exam boards to generate and use predictions collaboratively (note we have also required exam boards to work collaboratively to run the standardisation process, as outlined in Section 8.1). This means that a national prediction is generated for each subject, rather than exam board/specification specific predictions. This helps to ensure that students are treated the same regardless of their exam board. It also means that when evaluating the outcomes of exam boards' awards, we compare the outcomes for students matched to their prior attainment with the statistical predictions at the national level (rather than individual specification level). This would not be appropriate (or indeed possible) in a typical year given the different assessments offered by each exam board, and the likely differences in student performance.

5.3.1 The use of predictions at non-key grades

As outlined in Section 5.2, statistical predictions are typically only used to inform the setting of grade boundaries at key grades, and the remaining grade boundaries are calculated arithmetically. Given that there are no assessments this summer, the same approach is not possible. Further details of the approach to using predictions at the non-key grades is provided in Section 8, as part of the description of the standardisation model.

5.3.2 Adjustments in MFL

In recent years, we have performed extensive analysis of the grading standards in GCSE and A level modern foreign language (MFL) qualifications. This has included a significant body of work following concerns from stakeholders about the comparability of grading standards in French, German and Spanish with other subjects.[39] We have also considered the potential impact of native speakers on the maintenance of A level standards in French, German and Spanish.[40]

In 2017, we made an adjustment to the grading standards in A level French, German and Spanish (+1% at grade A with consequential increases taking place across the grade scale)[41] to reflect the impact of native speakers taking these subjects. In 2019, following the conclusion of our work on inter-subject comparability at A level, we committed to ensuring that the grading standards of these subjects did not become statistically more severe at grades A* and A. As such, we required exam boards to

39 Inter-subject comparability
40 Native speakers in A level modern foreign languages
41 Setting grade standards in A level modern foreign languages

ensure that, when using the statistical predictions to inform the setting of grade boundaries, outcomes were above predictions. While we intended to continue this requirement had exams being taken as normal this summer, this is not necessary given the way that statistical predictions are being used.

Our inter-subject comparability work has also considered the comparability of grading standards in GCSE French, German and Spanish with other subjects. We considered a wide range of evidence from different sources (for example, statistical evidence, stakeholder/exam board views and benchmarking against the Common European Framework of Reference for Languages) and we published the findings of our investigation in November 2019.[42] We concluded that the evidence represented a compelling case to adjust grading standards in GCSE French and German, but not in Spanish. We therefore committed to bringing the grading standards of GCSE French and German in line with those in Spanish. Because we have no statistical evidence and few specific stakeholder concerns about the relative difficulty of grades below grade 4, we determined that there was not a case to adjust grade 1 in any of the three subjects.

Following our announcement, we have been working with exam boards to consider how we should implement our decision. Our findings suggested that the adjustment needed was greater at grade 7 than at grade 4. We also said we may phase the implementation of the adjustment over two or more years, if a one-off adjustment would lead to undue unfairness between students in adjacent years of entry. We therefore agreed with exam boards that the national predictions should be adjusted as follows this summer:

- the predicted percentage of students achieving grade 4 and above should be increased by 1 percentage point
- the predicted percentage of students achieving grade 9 and grade 7 and above should be increased by 2 percentage points.

These adjustments are applied at the national level because of the approach to generating predictions this summer. We told schools and colleges making judgements about the CAGs in GCSE French and GCSE German not to try and take account of this adjustment, as it would be applied nationally as part of the standardisation process.

When we make adjustments to grade standards, we typically only make them to the key grades where statistical predictions are used in awarding – at GCSE these are grades 9, 7, 4 and 1. In a typical year, when an adjustment is made to one or more of the key grades, there is often an impact on outcomes at the non-key grades. For example, if an upwards adjustment in outcomes is sought at grade 4 leading to a lower grade boundary, then the calculated grade boundaries above and below grade 4 might also be lower than they would otherwise have been. This is not always the case, and it might not be the case for all grades, as it depends on the relative positions of the grade boundaries.

The standardisation approach being used this summer uses predictions at all grades (see Section 8 and Annex M for further discussion). This means that applying an adjustment to the key grades only would not impact on the outcomes at other grades. We therefore agreed, in consultation with exam boards, that an adjustment should be applied at the non-key grades to reflect the likely change in outcomes had

[42] Inter-subject comparability in GCSEs

exams been taken this summer. The size of the adjustments aligned with the need to make a greater adjustment at the top end of the grade range, and to make no adjustment at grade 1. As such, the following adjustments were applied to the national predictions: +2% at grades 9, 8 and 7, +1.67% at grade 6, +1.33% at grade 5, +1% at grade 4, +.67% at grade 3 and +0.33% at grade 2.

As we have said previously, we intend to keep under review the need for any further adjustment in future years to align grading standards in French and German with Spanish.

5.3.3 Adjustments due to the National Reference Test (NRT)

Another potential source of adjustment to standards in summer 2020 are the findings of the NRT. The details are not discussed here since, as usual, findings from the NRT and our decision on whether to make an adjustment in GCSE English language and/or mathematics will be published on GCSE results day, 20 August 2020.

5.3.4 Timing of predictive data

In a typical year, the outcome matrices that are used as the basis for generating predictions are created using students' grades from the time that results are issued in August. This means that the results upon which the statistical predictions are based do not account for any changes to grades that might occur post-results (for example, following a review of marking or moderation, or an appeal). This is necessary in a typical year because students' grades in August can only be based on their marks at this time (since any subsequent changes are unknown). This approach ensures a like-for-like statistical comparison for the purposes of setting grade boundaries. To do otherwise would lead to small increases in the statistical predictions each year (even if there were no changes in the prior attainment of the cohort), since the majority of post-results changes are in an upwards direction.[43] And, while such changes are small (and generally no greater than 1%), basing the predictions on the final grades that students received (rather that those as of August) would lead to year-on-year inflation that would not be fair to students in different academic years.

Given that students are not sitting any assessments this summer, it is not possible for there to be any post-results changes due to reviews of marking or moderation. And, while there is an appeals process, it is unlikely to lead to levels of change comparable to those in a typical year. This, therefore, risks potential disadvantage to this summer's cohort. Following consultation with exam boards we therefore decided that, to be fair to students, the statistical predictions should be based on the final grades awarded to students (using data from the appropriate reference series), rather than those awarded in August.[44] This means that any post-results changes

[43] Reviews of marking and moderation for GCSE, AS and A level: summer 2019 exam series

[44] An analysis of the differences in outcomes (due to reviews of marking and moderation) between when results are issued in August and when post-results closes is provided in Annex B, based on data from 2019.

that could occur in a typical year are accounted for in the statistical predictions. We specified this in our requirements[45] document that exam boards must comply with.

This issue also has implications for the historical centre data that is used in the standardisation model, as discussed in Section 8.

5.3.5 The role of absentees in predictions

Another issue to consider in relation to the statistical predictions is the role of absentees. In a typical year students taking a particular qualification can sit all of the required assessments or be partially or fully absent. Students that are partially absent have missed one or more of the required assessments (without a valid reason for being absent), while students that are fully absent have missed all of the assessments. Partially absent students are awarded a grade based on the assessments that they have sat, but it is not possible to estimate a grade for students that were fully absent.

Ordinarily, the outcome matrices upon which the statistical predictions are based are constructed excluding students that are partially absent for that qualification. And, when comparing statistical predictions with outcomes for students matched to their prior attainment in the current year (to inform the setting of grade boundaries), partial absentees are again excluded. This is to ensure that exam boards are comparing like-with-like, and that the relationship between prior attainment and outcomes is only based on those students that sat all of the required assessments. This is important since partial absentees are likely to perform differently[46], and it means that students are not advantaged or disadvantaged if the proportion of partially absent students changes from one year to the next.

In the context of awarding grades in summer 2020, no assessments have been taken so all students are in effect fully absent. Further, it cannot be known which students would have been partially (or indeed fully) absent had the assessments been available. This raises a question around whether partial absentees should be excluded from the outcome matrices upon which the predictions are based (as in a normal year), or whether they should be included.

We know from our analyses of data from previous years that the proportion of partially absent students (of the target age that would usually be included in the statistical predictions) is relatively low, but tends to differ between subjects and qualification level. For example, in 2019, the proportion of partial absentees was generally higher at GCSE than A level for 16- and 18-year-olds (around 0.5% and 0.2% of the total entries, respectively – see Annex C for a breakdown by subject and qualification). This suggests that the impact of choosing one approach over the other is likely to be relatively small. Nonetheless, we are mindful of ensuring that our approach does not disadvantage students given the unique circumstances this summer.

Following consultation with exam boards, we therefore determined that the outcome matrices should exclude partial absentees (as in previous years). This means that the statistical predictions are only based on those students that sat all of their assessments in previous years, and we are therefore predicting outcomes this

[45] Requirements for the calculation of results in summer 2020

[46] Chamberlain, S. (2008) Research in Education, 79(1), 53-66
Sleeping in or Selecting Out?: Candidates' Absence from GCSE Examinations

summer based on the assumption that all students would have sat all of their assessments (had they been available). This is unlikely to have been the case, but our approach is in keeping with the desire to give students the benefit of the doubt this summer. The alternative would have resulted in slightly lower predicted outcomes and would likely have resulted in a small degree of disadvantage being spread across all students, including those that would have sat all their assessments had they been available. This is because we cannot distinguish between those students that would have sat all of their assessments this summer, and those that would have been partially absent.

In addition to some students being partially absent in a typical year, there are a small number of students that are completely absent and fail to sit any assessments in the qualification that they are entered for. These numbers are generally low (for example in 2019, around 0.8% of GCSE and 0.2% of A level entries were fully absent), and their outcomes are not routinely included in the outcome matrices used to generate statistical predictions. Similar to partial absentees, it is not possible to determine which students would have been fully absent this summer, so it is assumed that all students with entries would have sat all of their assessments.

6 Potential approaches to standardisation

6.1 High-level categorisation of approaches

At the highest level there are 3 approaches to the standardisation of CAGs this summer. These approaches are:

- **macro-level standardisation** where the adjustment applied is defined by a population level relationship that is applied to the whole cohort
- **meso-level standardisation** where centre-level statistical estimates are used to standardise each centre
- **micro-level standardisation** where estimates are formed based on the characteristics of individual students

These different approaches are described below.

6.1.1 Macro-standardisation

The simplest approach to standardising centres' results this summer would be to apply a single common transformation to the CAGs submitted by centres entering for a qualification. These approaches would assume either that all centres should follow the same value-added relationship or that the relative attainment between centres is contained already within the CAGs and should be relied upon and retained (albeit shifted towards overall leniency or severity).

The most straightforward macro approach would be to apply a linear transformation to the CAGs, thus retaining the inter-centre relationship they contain. This would apply a linear shift, upwards or downwards, as necessary, with the sole aim of maintaining overall national standards. As with any other approach, it would be necessary to determine how fine adjustments are applied with students being prioritised for promotion or demotion to the grade above or below based on the rank orders. While this approach would achieve outcomes that match those of previous years at the cohort level, this would not reflect the inevitable differences in approach taken by different centres to produce the CAGs. Assuming an overall level of optimism in the CAGs, this approach would reward students at centres whose CAGs contained the greatest levels of optimism (as their final outcomes would retain much of this generosity) and penalise students from centres who (correctly or incorrectly) took a more cautious approach to producing their CAGs (whose outcomes would remain towards the bottom of the distribution). As this approach would fail to address inter-centre differences in approach, it was deemed neither fair nor credible.

A slightly more complex approach, but still one reliant on macro relationships would be to make global assumptions about the value-added relationship demonstrated by centres. This would be operationalised by assuming that all centres entering students for a subject would likely demonstrate the same value-added relationship between their students' prior attainment and the grades they should receive. The same overall relationship – usually applied at cohort level – could be applied to each centre to estimate their results for the current year.

Practically, this would be problematic for centres with students who do not have the relevant measures of prior-attainment. However, setting this aside, it is widely recognised that different centres exhibit different value-added relationships and such an approach would fail to recognise these legitimate centre-level differences. Any sense that this approach would fairly standardise centres therefore appears to be flawed.

Since any macro approach would fail to sufficiently reflect the available evidence relating to individual centres and would, therefore, fail to address inter-centre differences in a credible way, this category of approaches was deemed inappropriate to be pursued for standardisation.

6.1.2 Micro-standardisation

An alternative approach would be to operate at the other end of the scale in terms of granularity and take an approach where students' results are predicted at the individual level. Such an approach would rely on the individual statistical indicators, such as prior-attainment.

While the use of measures of prior-attainment are commonplace in research studies seeking to control for differences in the underlying ability of students/participants,[47] and also are routinely applied for operational predictive purposes by exam boards (as described in Section 5), they are only sufficiently meaningful to be informative for groups of students. For example, an individual student with a high prior attainment may be more likely to achieve a higher grade in a subject compared with a student with a lower prior attainment, however, it would be wrong for this to predetermine the results for individuals on this basis on a student-by-student level.

Typical correlations between prior attainment and attainment in a GCSE are 0.34 to 0.76[48] and 0.57 to 0.71 at A level[49]. These provide a valuable indicator when applied to groups but would be an over-interpretation of the measures when seeking to predict the outcome for an individual student.

Taking such an approach would also close off, or certainly undermine, the value in the rank orders provided by teachers as described in Section 2.3. This evidence is being collected to provide important information about the abilities of students that cannot be sufficiently captured by purely statistical means. Putting greater reliance on statistical measures relating to an individual student would, therefore, risk disordering the rank order submitted by the centre in an indefensible, and likely invalid, way. In summary, to be of value, individual predictions would have to be allowed to override teacher rank orders, but they would not achieve a level of reliability that would warrant them being used in this way.

[47] Pinot de Moira, A., Meadows, M.L. & Baird, J-A. (2019) The SES equity gap and the reform from modular to linear GCSE mathematics, *British Educational Research Journal*, https://doi.org/10.1002/berj.3585.

[48] Benton, T. and Sutch, T. (2013). Exploring the value of GCSE prediction matrices based upon attainment at Key Stage 2. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment

[49] Benton, T. and Bramley, T. (2017). Some thoughts on the 'Comparative Progression Analysis' method for investigating inter-subject comparability. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment. Retrieved from: https://www.cambridgeassessment.org.uk/Images/416591-some-thoughts-on-the-comparative-progression-analysis-method-for-investigating-inter-subject-comparability.pdf

Further, such an approach that would override teachers' rank ordering of students, might be considered akin to profiling; making individual decisions about the results of students based purely on historical data, rather than drawing on any evidence of human judgements or evidence relating to their abilities in the qualifications being awarded.

6.1.3 Meso-standardisation

The final group of possible approaches are those referred to here as meso-standardisation. These methods fall between the macro and micro approaches discussed above and seek to strike a balance between addressing inter-centre differences in approach to determining CAGs (something which macro approaches fail to do effectively) and not over-interpreting statistical indicators for individual students (as in the micro approaches). Their aim is to identify relevant information relating to each centre in each subject to determine the most appropriate outcomes based on the statistical evidence available and relying on those centre-level estimates for standardisation.

This group of approaches is likely to be more complex than both the macro- and micro-standardisation approaches, but this cost would come with a likely improvement in predictive accuracy and fairness. This group of approaches situate the most likely outcomes for an individual centre based on the statistical evidence available, thus achieving an appropriate inter-centre alignment and then draw on the rank order information submitted by centres to determine the individual student grades. These approaches assume correctness of the rank order provided by the centre due to the teacher estimate being the most reliable 'fine' student measure available.

Given the potential for macro approaches to fail to reflect legitimate differences in centre characteristics and the risk of micro approaches significantly over-interpreting student level estimates, it was decided that meso-standardisation approaches would form the basis of further exploration.

6.2 Description of the potential models

The information available for the purposes of standardisation this summer are:

- the historical performance of students entering for each subject in each centre
- the prior-attainment relating to students in the historical data-set
- the prior-attainment of students who were due to sit their assessments this summer
- the CAGs and rank orders submitted by centres

A range of different approaches to realising meso-standardisation were explored, all following the same two-step process:

> Step 1) locate the centre in terms of their expected distribution of results based on an historical statistical model.
>
> Step 2) populate that expected distribution of results based on the teacher estimates.

The primary difference between the approaches considered here is in step 1: the approach used to determine how centres' results are located and distributed.

It should be noted that the models for locating centres are based purely on historical statistical information (i.e. previous centre performance and previous and current student prior attainment).

The methods considered to perform step 1 of the process can be considered in three different groups:

1. Mark-based regression.
2. Grade-based regression.
3. Direct centre-level performance.

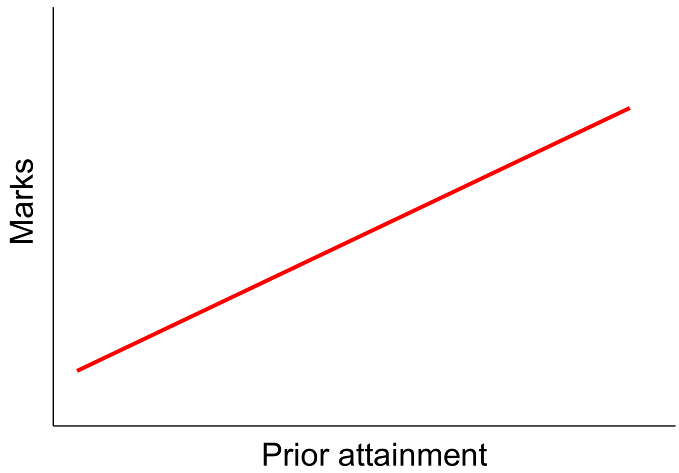
These different approaches are described below.

6.2.1 Mark-based regression (MBR)

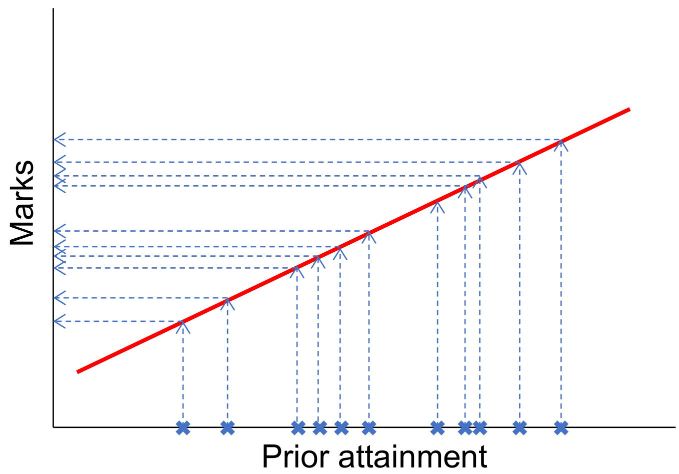
The first category of approaches takes advantage of the finest grained information available in the historical data sets - individual students' marks. They rely on generating a predictive model based on historical data which then takes the information available about students entering for a qualification this year and generates a prediction of the results they would have achieved, had they had the opportunity to sit their exams. The predictive models in this group of approaches are based on the historical relationship between students' prior-attainment and the final marks they achieved in the qualification.

Before describing the detailed considerations necessary to implement such an approach, a high-level description is outlined. The steps required for a mark-based regression approach would be as follows:

- Step 1 – Fit a statistical model, based on historical results in the subject. This model is based on the relationship between students' prior attainment and the mark they achieved in the qualification. The simplest form of this relationship is shown below.



- Step 2 – For each student entering the qualification this summer, determine their value of prior attainment.
- Step 3 - Based on the predictive model defined in Step1 use the prior-attainment information for students identified in Step 2 to generate a predicted mark for each student.



- Step 4 – Bring together the student-level marks for each centre to form a centre-level mark distribution. Then, reassign the modelled marks to students based on the rank order provided by the centre rather than that which results from the predictive model.

From model

Modelled uniform marks

Candidate	Prior Attainment	Modelled mark
Candidate 1	7.4	843
Candidate 2	6.4	642
Candidate 3	4.8	411
Candidate 4	3.4	303
Candidate 5	2.9	261

Order/mark allocation revised based on teacher rank order

Final allocation of marks

Candidate	Rank	Modelled mark
Candidate 1	1	843
Candidate 3	2	642
Candidate 4	3	411
Candidate 2	4	303
Candidate 5	5	261

- Step 5 – produce a national mark distribution for prior-attainment matched students by bringing together marks from across all centres.
- Step 6 – using the mark distribution produced in Step 5, set cut-scores at marks that best achieve the subject-level prediction as described in Section 5. This stage of the process is analogous to the preparation performed by exam boards in a typical year to produce statistically recommended grade boundaries.
- Step 7 – award grades to students based on the marks allocated to them based on the rank order in Step 4 in relation to the cut-scores.

It can be seen from this description that, while this approach relies on statistical predictions for individual students, the risks of over-interpreting the relationship between prior-attainment and result on the individual outcomes (which would be the case for micro-standardisation) are avoided. This is achieved by combining the student-level information to centre-level before then allocating grades to individual students based on the more reliable rank order provided by the centre.

The use of marks as the basis for this prediction provides the key advantage that it allows a direct mechanism through which to ensure overall outcomes in a qualification are maintained, as described in Step 6.

This high-level description has overlooked a number of important issues in applying the approach as discussed below.

Predictive model

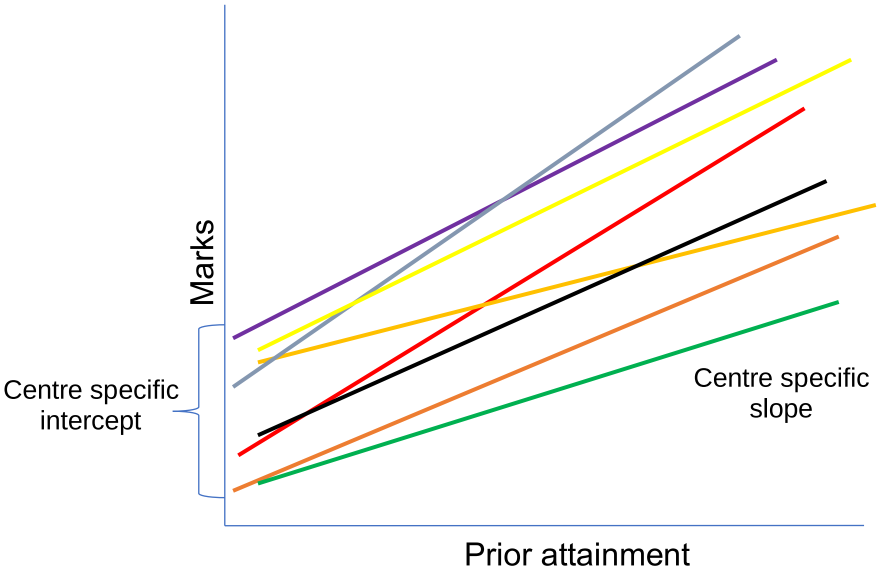
For simplicity, the procedure outlined above referenced a single relationship between prior-attainment and results that was applied to all students of the simple form:

$$Y_i = \beta_0 + \beta_1 A_i$$

where Y_i is the mark achieved by student i with prior attainment measure A_i. This predictive model was fitted at Step 1 and then used to generate predictions at Step 3.

As discussed in Section 6.1.1, it is recognised that different centres demonstrate different rates of value-added. It would, therefore, be inappropriate to use such a simple single relationship to predict the behaviour of all centres. There is, however, an overall relationship for the cohort between measures of prior-attainment and the final mark that they achieve. In these circumstances where it is necessary to reflect the clustering of data, but also to draw on the statistical power provided by a population-level relationship, multi-level models[50,51] can be used. These approaches facilitate the production of individual centre-level estimates reflecting local relationships to inform the student level estimates, but in a more reliable way than would be possible if attempting to analyse each centre in isolation.

An example of the sort of relationship that may be established through fitting such a model is shown illustratively in the figure below. In this model, each different coloured line represents the relationship between prior-attainment and the marks achieved for each individual centre.



50 Goldstein, H. (2011). *Multilevel statistical models (4th edition)*. Chichester: John Wiley & Sons.

51 Snijders, T. & Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modelling*. London: Sage Publications.

In this instance, the model allows each centre's relationship to vary in terms of the gradient of the line and where it intercepts the 'marks' axis. However, a range of different forms of this model were considered for use. These included random intercept models of the form:

$$Y_{ij} = \beta_0 + \beta_1 A_{ij} + \gamma C_j + u_j + e_{ij}$$

Random slope models of the form:

$$Y_{ij} = \beta_0 + \beta_1 A_{ij} + \gamma C_j + u_{0j} + u_{1j} A_{ij} + e_{ij}$$

And polynomial forms:

$$Y_{ij} = \beta_0 + \beta_1 A_{ij} + \beta_2 {A_{ij}}^2 + \beta_3 {A_{ij}}^3 + \gamma C_j + u_j + e_{ij}$$

where:

- Y_{ij} is the uniform mark achieved by student i in centre j
- A_{ij} is the student level measure of prior attainment
- C_j is a summarising measure(s) of centre level historical performance (articulated as mean uniform mark, mean grade or cumulative percentage outcomes at key grades)
- u are the centre level random effects
- β_0, β_1, γ and δ are the fitted regression coefficients
- e_{ij} is the student level residual

During preliminary test testing of these different approaches, it was clear that, applying one of the more complex random slope or polynomial forms of the model was neither productive nor desirable. For the random slope models the computational overhead was significant with the fit failing to converge in a number of cases. The lack of convergence is problematic in the context of needing to apply the standardisation approach across 157 subjects (each requiring their own model) and the desirability of applying the same approach as broadly as possible. Regarding the polynomial models, the increase in variance explained was extremely marginal. For example, when fitting the models for A level biology, the linear variants resulted in R^2 values in the range 0.54-0.55 and, for the equivalent polynomial forms, the values were in the range 0.55-0.56. This increased complexity was, therefore, not considered to be necessary.

Marks and mark scale

Fundamental to this approach is access to the historical marks of students. This information is routinely available to exam boards through their operational systems. The basis on which these marks are defined, however, introduce three challenges.

First are the differences between mark scales used by different exam boards for different specifications in the same subject. As will be discussed in Section 8, it is important that data are combined at the national level for the purposes of the standardisation process this year. Different maximum marks used by different exam boards in the same subject makes it less easy to use marks across different specifications.

The second issue is the variation in grade boundaries over time. In every typical exam series, exam boards set grade boundaries on their assessments to ensure the maintenance of grading standards over time. The predominant cause of grade boundaries having different positions between years are variations in the difficulty of the assessments making up that qualification. By definition, this indicates that raw marks should not be considered to be equivalent over time and, therefore, an approach which relies on this being the case would be limited in its accuracy and may be biased towards centres previously entering for a subject with a particular exam board or with different amounts of historical data in different years.

The final challenge is performing this activity immediately following the qualifications reform process (see Section 2). More specifically, not only will the length and construction of the mark scales have changed through the process of reforming the qualifications but, in the case of GCSE, the grade scale has also changed from being defined A*-G to 9-1. This is problematic if seeking to draw historical data from across years which may span the transition from one version of a specification to another or draw data from a single year in which both legacy and reformed versions of a qualification were available.

To overcome these problems, the mark-based regression approaches transform the subject level raw marks onto a uniform universal scale that can be applied across specifications in the same subject and over time to account for differences in grade boundary position and structural changes. This enables all specifications from across exam boards, over time and within a subject to be combined. While differing in use to the uniform mark scale (UMS) used in the operationalisation of modular qualifications[52], the scale used here can be considered analogous with such an approach.

The maximum uniform marks for each subject are, arbitrarily, set to 1000. The three proposed mapping schemes used are presented below.

A level:

Grade	Corresponding uniform mark at boundary mark
A*	900
A	800
B	700
C	600
D	500
E	400

52 Guide to the Uniform mark scale (UMS)

Reformed GCSE:

Grade	Corresponding uniform mark at boundary mark
9	900
8	800
7	700
6	600
5	500
4	400
3	300
2	200
1	100

Legacy GCSE:

Grade	Corresponding uniform mark at boundary mark
A*	850
A	700
B	550
C	400
D	325
E	250
F	175
G	100

Handling of missing data

The role of prior-attainment is clearly important as the independent variable on which the student-level prediction is based when using the mark-based regression approaches. In many instances, however, students may not have these measures of prior-attainment available leading to an issue with missing data. (A table of prior-attainment match rates can be found in Annex D.) This issue could be resolved by imputing either the value of missing prior-attainment, the mark or the final grade. The following approaches were investigated:

i. **Post-hoc-slotting** This approach addresses the issue of missing data at the end of the process by determining students' grades based on other students in the centre for whom data is not missing. In many cases, where the number of students within a centre without prior attainment are in the minority, it is relatively trivial to use the rank order provided by teachers to identify the appropriate grade for a student who does not have a predicted grade generated through the process outlined above. Where an unmatched student falls between two others with the same grade, they can be allocated the same grade as those surrounding them. In instances where the students either side have different grades, the decision whether to allocate the higher or lower grade is less straightforward. However, this could be resolved by moving in the direction of the student's CAG. It should be noted that this approach requires that students with no prior attainment data are explicitly excluded from the predictive model until this final stage.

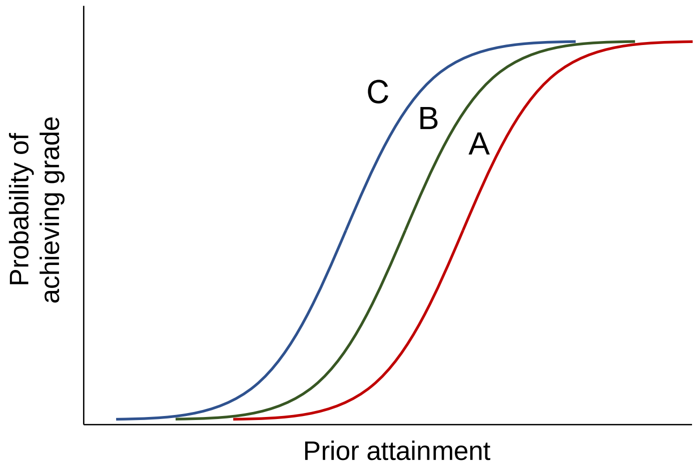
ii. **Imputation based on available prior attainment distributions** This approach is intended to explicitly deal with missing prior attainment data by imputing a measure based on other existing information. Where the proportion of students with missing data within a centre is too great for the estimation process to work effectively the process of post-hoc slotting is likely to be ineffective and inappropriate. This approach to imputation relies on the matching of centres without prior-attainment data to those with prior-attainment based on characteristics such as historical centre performance and centre type, Here, missing values are replaced with values of prior attainment randomly extracted from a distribution with parameters matched to the empirical distribution of prior-attainment for similar centres. In addition to being technically complex, this type of approach is challenging to defend due to the reliance of many imputation techniques on drawing random samples from distributions that may impact on the grades for individual students.

iii. **Assume similar performance to other unmatched students** An alternative approach is to treat prior-attainment as a categorical rather than continuous variable. Using this approach, a category can be attributed to those without the relevant information assuming this group would perform similarly to previous unmatched students from the centre. As treatment of prior-attainment in this way would lead to multiple students being attributed the same mark in the centre, the distribution of marks can then be determined by sampling the estimated from the relevant error distribution.

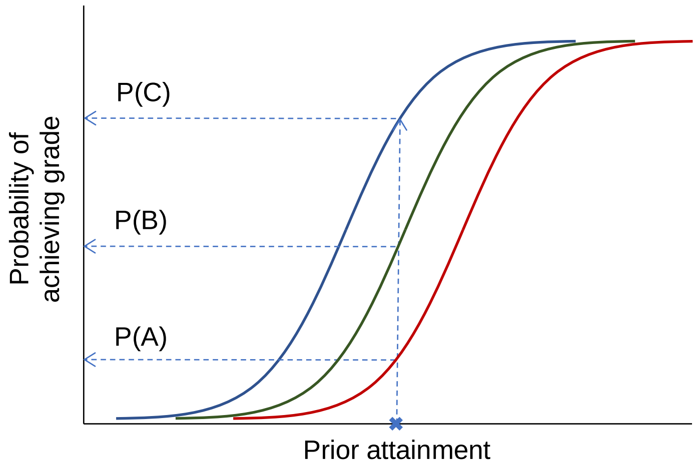
6.2.2 Grade-based regression (GBR)

A similar approach to that described above can be implemented using grade-based rather than mark-based models. The process followed for this category of approaches is similar to that outlined in Section 6.2.1. with two key differences: the form of the model (Step 1 of the process) and how the student level estimates are brought together to form the centre-level prediction. The process for grade-based regression approaches was as follows:

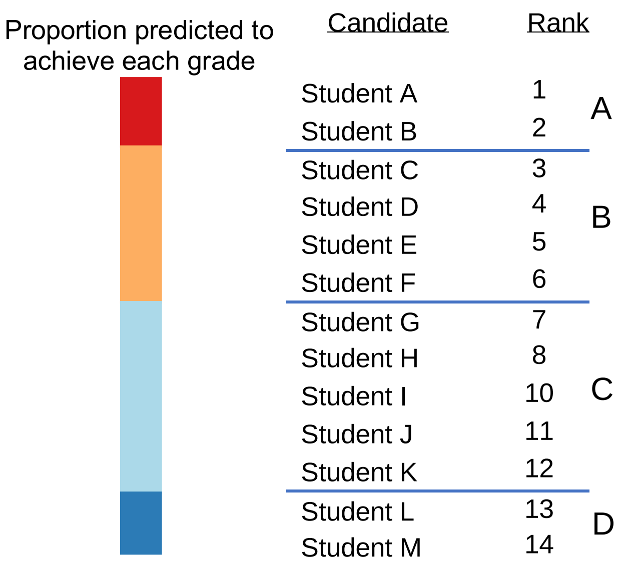
- Step 1 – Fit a statistical model, based on historical results in the subject. This model is based on the relationship between students' prior attainment and probability of achieving each grade (or higher) in the qualification.



- Step 2 – For each student entering the qualification this summer, determine their value of prior attainment.
- Step 3 – Based on the predictive model defined in Step1 use the prior-attainment information for students identified in Step 2 to generate the probability of each student achieving each grade.



- Step 4 – Aggregate the student level probabilities at each grade to determine the overall predicted grade distribution for the centre in the subject.
- Step 5 – Assign grades to students within the centre for the subject, based on the rank order of students provided by the centre and the proportion of grades predicted for the centre in that subject. Students are allocated to meet the centre level prediction for the subject as closely as possible.



Having completed these steps, students will have been assigned notional grades, however, as discussed in Section 5, it is necessary to have the facility to adjust overall outcomes. This is something that is naturally accommodated in the MBR approaches due to the facility to set grade boundaries. For these approaches it was necessary to impute marks for individual candidates. As this element of the process that was necessary in the final implementation, a detailed description of the approach is provided in Section 8.2

As with the mark-based regression approaches, the representation above is a simplification as the relationship shown at Step 1 has been shown here to be a single relationship for all centres. In practice, it would be unreasonable to make this assumption. For the purposes of the analysis that follows, both ordinal and simple multi-level logistic models were considered to reflect the structure of the data. Resolution of missing data can also be approached in a similar way to that outlined above.

6.2.3 Direct Centre-level Performance (DCP)

The group of approaches that are the most intuitive are those being termed direct centre-level performance approaches. As the name suggests, the starting point for these methods is to directly use previous years of performance of the centre in the subject to form the basis of the prediction.

The basis of the approach is to assume that a centre will perform the same in a subject this year as they have across recent years, taking into account any changes in underlying ability of students. The changes in ability of the cohort are based on a comparison of the prior-attainment of students this year compared to the prior-attainment of students in the historical cohort.

At a high level, this is achieved by carrying out the following procedure:

> Step 1 – Identify students in the historical data who have sat a qualification in the subject with that centre. Determine the grade distribution achieved by those historical students.
>
> Step 2 – Based on the national historical value-added relationship between prior-attainment and grade achieved, generate a prediction for how that historical cohort would have been expected perform.
>
> Step 3 – For each student entering the qualification this summer, determine their value of prior-attainment.
>
> Step 4 – Based on the national historical value-added relationship between prior-attainment and grade achieved, generate a prediction for how this year's cohort would have been expected to perform.
>
> Step 5 – Based on the difference between the predicted outcomes for the current students (calculated in Step 4) and the predicted outcomes for the historical students (calculated in Step 2), adjust the historical grade distribution. This adjusted grade distribution defines the outcomes for the centre in the subject this summer.
>
> Step 6 – Assign grades to students within the centre for the subject, based on the rank order of students provided by the centre and the proportion of grades predicted for the centre in that subject. Students are allocated to meet the centre level prediction for the subject as closely as possible. This is identical to the approach taken in Step 5 of the grade-based regression approaches as illustrated above.

As with the other methods described, this approach uses the prior-attainment of students to reflect the changes in ability profile of the cohort over time. However, in contrast to the regression approaches, students without prior-attainment can be handled implicitly in the approach (although this is closely analogous to iii. on page 44). As outlined above, the principle of the approach is to adjust the predicted grade distribution of the centre in the subject, where the evidence in the relative prior-attainment profile over time indicates that is appropriate. In the absence of prior-attainment for either the current year's cohort or in previous years, it is assumed that the historical performance of the centre in previous years will be carried forward to this year. In many cases, centres will have a mix of prior-attainment matched students and those without prior attainment. The extent of the adjustment described in Step 5, due to differences in the prior-attainment cohort of students, is weighted by the proportion of prior-attainment matched students across the historical and current data.

With these approaches (as with the GBR approaches), it is necessary to provide the facility to perform adjustments to outcomes at the cohort-level at the end of the process. Again, this applies the approach to imputing marks as described in Section 8.2.8.

This section has provided a high-level view of the three categories of approach that were explored to standardise CAGs this summer. The section that follows outlines the testing procedure and results leading to the selection of the final approach to be implemented.

7 Model testing and results

7.1 Approach to testing

One of the challenges of evaluating the range of approaches summarised in the previous section is identifying a basis for validation. Results data for 2020 against which to compare predictive performance are clearly not available. To provide a basis for comparison it was, therefore, necessary to apply the standardisation approaches to the task of predicting students' grades in 2019. The actual data from 2019 then provided a set of results against which the outputs of the modelling could be compared. This is illustrated in Figure 7.1, assuming the use of 3 years of historical data.

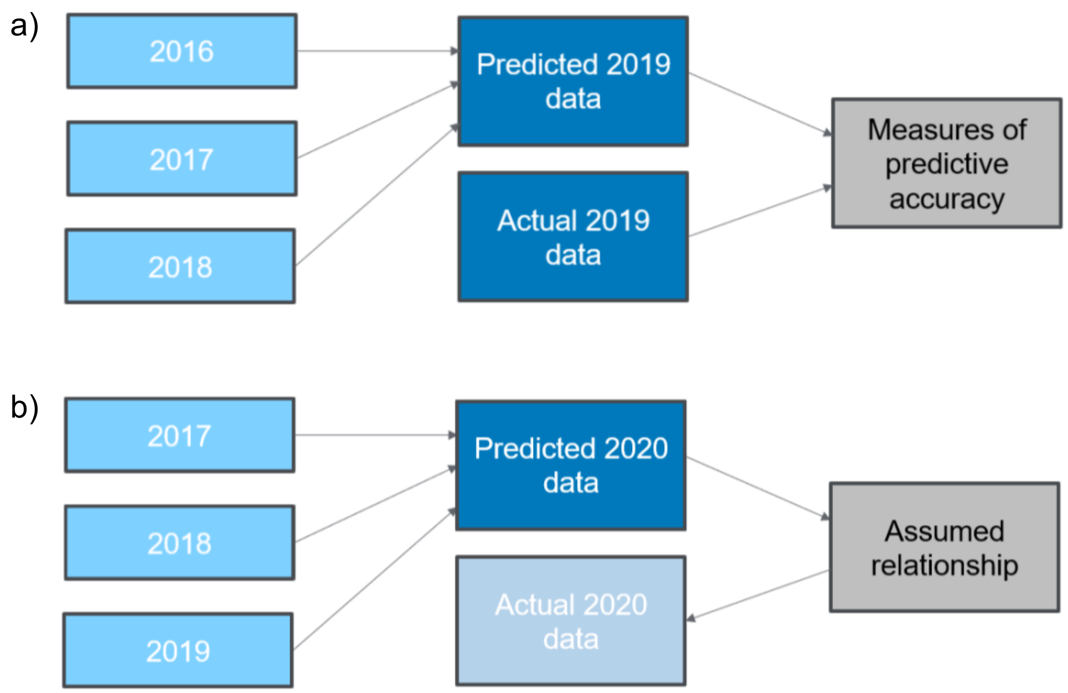


Figure 7.1 a) Use of 2019 actual data as a basis of comparison for evaluative purposes and b) the assumed relationship in 2020

As can be seen from Figure 7.1, to provide an authentic test of the different approaches, the years of historical data must also shift accordingly. Despite data up to 2019 being available for the purposes of prediction in summer 2020, using 2019 data for the purposes of evaluating the predictive performance in the same year would be inappropriate. This would risk overestimation of (or at least cast doubt upon) the predictive performance indicated by the test results.

While this testing approach is necessary, it does have limitations due to the timing and phasing of qualification reform, with some of the approaches being considered being dependent on the continuity of grading scale over time. For GCSE, therefore, for each subject, it was not possible to replicate the number of post-reform years that would be available for standardisation this summer. Based on the analysis presented below, this may lead to an underestimate of the predictive accuracy in some cases.

Where, through the process of testing, it was necessary or desirable to draw on rank orders of students, the actual rank order within the centre based on the marks achieved in 2019 were used as a replacement. The only limitation of this approach is that, unlike the rank orders collected from centres in summer 2020, this would lead to ties due to some students in the centre scoring the same marks as one another.

To evaluate the performance of the different standardisation approaches it is necessary to define some metrics that can be used as a basis for comparison. The key metric used was the probability of the predicted grade matching the grade actually achieved by the student. The predictive accuracy measures reported here, therefore, represent the rate with which there is an exact match between the predicted graded and the actual grade for each individual student. To provide a realistic comparison of the approaches, the methods described in Section 6 were applied such that the model output replicated the appropriate overall cohort-level outcome for the subject.

Comparisons between the different approaches presented in Section 7.3 are based on results produced by Ofqual and the exam boards. To ensure a common basis for testing, a single national data set was generated and distributed to each organisation.

7.2 Number of years of historical data

As referenced above, one design decision that is common to all of the standardisation approaches regards the number of years of historical data to be used as a basis for prediction. There is an intuitive trade off here that can be tested empirically; the inclusion of more years of data to provide a larger number of students in the statistical basis for the prediction, and the inclusion of fewer years to improve the recency of the data. It should be noted that, in live implementation, special arrangements were put in place to handle instances where the number of students in the historical data (or indeed the current year of data) were particularly small (see Section 8.4), however, these approaches where not implemented during the testing phase meaning the predictive accuracy metrics presented here are likely to be pessimistic in some instances.

While it would have been desirable to fully test all variants of all approaches using a range of historical years of data, it was not practicable to do so. Moreover, it was desirable to provide more information to teachers regarding the data that would be used by the statistical model as early as possible in the process. This was valuable for centres seeking to take a data-led approach to establishing their distribution of CAGs to ensure that the grades they submitted to the process were likely to be appropriate.

To support these groups of teachers, it was necessary to take the decision over the number of years of historical data before the selection of the final model had taken place. The two primary models being explored at the time were the core DCP approach and the basic mark-based regression model. The predictive accuracy of these two approaches with differing numbers of years of historical data were compared for a selection of subjects.

The relative predictive accuracy measures, compared to the use of a single year of data, are shown in Figure 7.2 in terms of percentage point difference. The black lines are from applying the DCP approach and the blue lines are from the MBR approach.

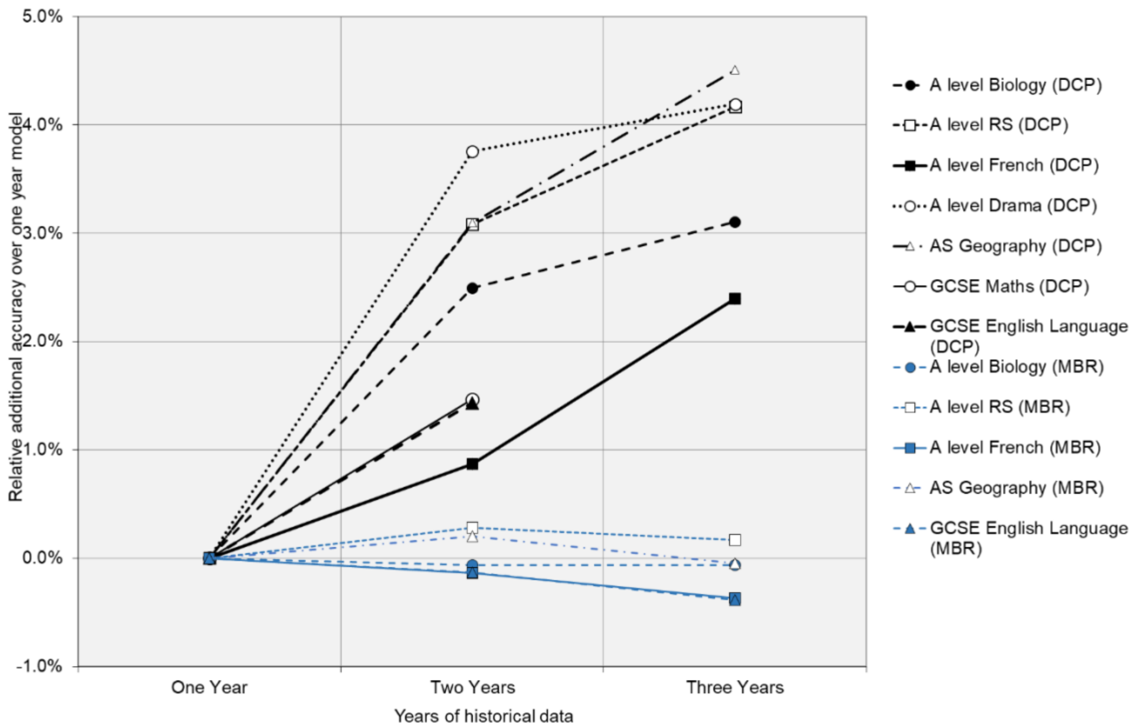


Figure 7.2 Relative predictive accuracy for a DCP (black lines) and MBR approach (blue lines) based on 1, 2 or 3 years of historical data.

This plot shows, for the DCP approach, the predictive accuracy increases with the inclusion of 2 years of historical data compared to 1 with a further increase (albeit typically smaller) when including 3 years of data compared to 2. In contrast, the MBR approach showed mixed results with no apparent systematic variation in predictive accuracy as the number of historical years of data increased.

Based on this evidence and the early indication of the better predictive performance of the DCP approach, it was decided that the following years of historical data would be used in the predictive models:

- AS/A level – three years of data
- GCSE – two years of data for reformed specifications first awarded in 2017 and 2018 and a single year for those first awarded in 2019.

The rationale for the selection of this time period for AS and A level was the improvement in accuracy, demonstrated above, in combination with the ability to consider data across the transition from the legacy to the reformed specifications because of the commonality of the grading scale over time.

While making this decision, however, it was important to consider the potential unintended consequences of relying on data drawn from across the transition between pre- and post-reform qualifications. Issues may have occurred with this approach if there had been a significant change in the overall outcomes for qualifications across the transition to the new specifications due to, for example, to unfamiliarity of teachers with the material. However, as described in Section 5, the use of statistical predictions to support awarding to protect the results achieved by students at the point of transition to the new specification has ensured that there was no cohort-level impact.

Other issues may also have arisen were there to have been a significant shift in the rank order of centres through the transition. In the lead up to the first award of the reformed qualifications, we published cautionary messages regarding the potential for there to be greater levels of instability in centre-level outcomes due to the varying effectiveness with which centres may have transitioned to the new specifications[53]. The potential volatility in outcomes did not, however, transpire.[54] These factors, combined with the improvement in accuracy for the DCP approach when applied across this timeframe, suggested that these potential issues have not adversely impacted the performance of the approach.

The rationale for using two years of historical data at GCSE (where available) is the ~1.5% point increase in predictive accuracy by using the additional year of historical data. However, due to the timing of reform, it was only possible to test the DCP approach across 2 years for those reformed subjects awarded for the first time in 2017. It would therefore, have been speculative to have extended this time frame to 3 years without being able to test it for these qualifications. This would also have also become problematic for centres taking a data-led approach to producing their CAGs due to the change in grading scale through qualification reform.

This information was published in Ofqual's blog on 15 May[55]. All the remaining analyses presented here use the number of years of historical data as specified above.

Because of the need perform testing on 2019 data and the recency of the reforms, in most cases, this reduced the number of years of data that were available for the testing of GCSE qualifications. This is likely to slightly underestimate the predictive accuracy for most of these subjects due to only one year of post-reform data being available for predicting 2019. This effect will not be a limitation for those subjects when running the model live in 2020, however, it will not be possible to measure this difference.

53 Letter to headteachers: summer 2017 exams and awarding
54 Variability in A level results for schools and colleges 2016-2018
55 Making grades as fair as they can be: advice for schools and colleges

7.3 Test results

As highlighted in Section 6 there are a range of variants that could be applied for the MBR, GBR and DCP approaches. The restriction on the number of years of historical data, as described above, reduces the number of possible configurations it was necessary to test fully. Outlined in Table 7.1 below are the variants of the approaches that were prioritised for deeper exploration.

Table 7.1 Summary of approached identified for further exploration

Approach	Category	Description
Approach 1 – Standard DCP approach with the estimation of imputed marks based on rank order	DCP	This approach employed the procedure described in Section 6.2.3 with the extension of imputing students' marks based on a combination of the student rank order and the centre level prediction.
Approach 2 – Mark based DCP with distributional mark assumptions	DCP	This approach is a mark-based equivalent of Approach 1 that operates by carrying forward the mean and standard deviation of centres' (adjusted) uniform mark distributions from previous years.
Approach 3 – Modified DCP approach	DCP	This approach is identical to Approach 1 with the variation that the weight put on the historical data is reduced as the size of the historical entry reduces.
Approach 4 – MBR with centre-level value-added only	MBR	The model in this approach allowed centre-level estimates of value-added to vary without specifying additional centre-level variables. Marks were imputed for students missing prior-attainment based on post-hoc slotting and interpolation.
Approach 5 – MBR with pooled centre-level performance variables	MBR	As Approach 4 with the addition of a centre-level historical mean grade variable. Marks were imputed for missing students based on post-hoc slotting and interpolation.
Approach 6 – MBR with imputation based on the national distribution for unmatched students	MBR	Specification of prior-attainment in the model was categorical, including a category for students without prior-attainment to allow prediction of this group. Mean centre-level prior attainment was specified as a centre-level variable.
Approach 7 – MBR with marks allocated using distributional summary	MBR	As Approach 6 with student-level marks determined based on a predicted mean and variance of marks for centre, allocated based on rank order position.
Approach 8 – MBR using penalised optimisation	MBR	Penalised constrained optimisation of a random slope model. Students without prior-attainment had this value imputed

		based on the average prior-attainment for students with the matching CAG[56].
Approach 9 – Multi-level logistic regression	GBR	This approach following the method outlined in section 6.2.2. using a multilevel logistic model
Approach 10 – Multi-level ordinal logistic regression	GBR	As approach 9 with an ordinal model
Approach 11 – Simple logistic regression	GBR	As approach 9 removing the hierarchical structure of the data and applying a simple logistic regression

Plotted below are the overall accuracy metrics for the 11 approaches outlined above. Figure 7.3 includes four A level subjects that were selected to provide a range of entry sizes and subject types. These subjects are biology, French, drama and religious studies. An equivalent plot is provided in Figure 7.4 for the following GCSE subjects: English language, mathematics, history and music. The values included in these plots are tabulated in Annex E. It should be noted that Approach 10 (multi-level ordinal logistic regression) is omitted from all GCSE analyses due to issues with the convergence of the models. This ruled Approach 10 out as a potential solution at this point, however, the results are included for A level, where possible, as a basis for comparison.

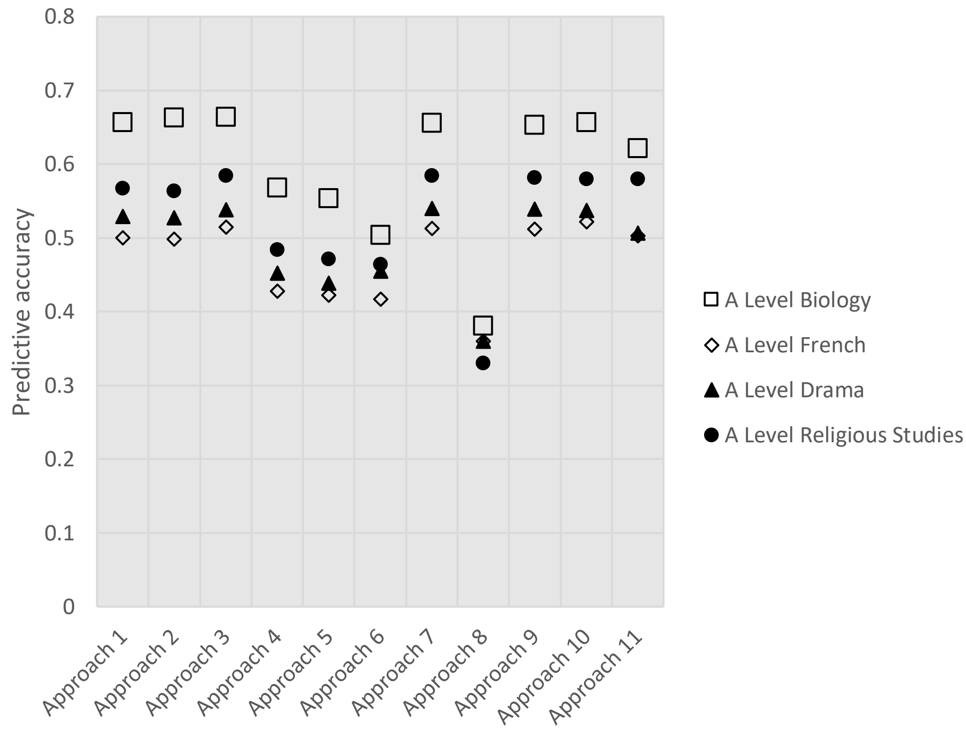


Figure 7.3 Overall predictive accuracy for the different models for A level biology, French, drama and religious studies

56 For the purposes of testing, students actual grades achieved were used for the purposes of imputation.

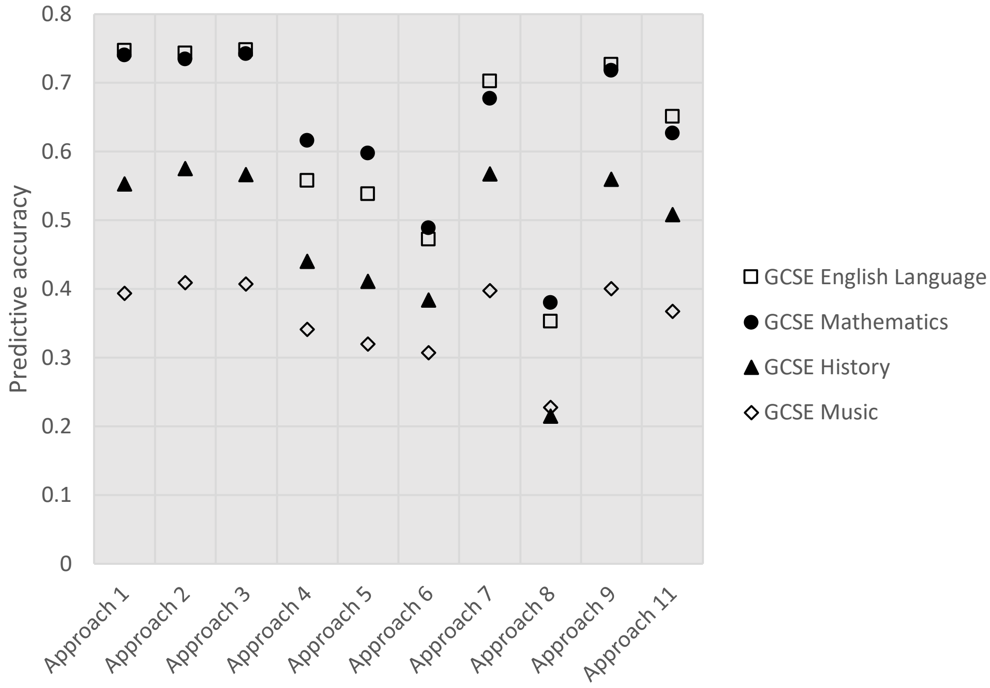


Figure 7.4 Overall predictive accuracy for the different models for GCSE English language, mathematics, history and music

From an initial review of the results, these figures show that all three of the DCP approaches (Approaches 1 to 3) perform well compared to the other approaches. There is greater variability across the MBR and GBR approaches with Approaches 7 and Approach 9 (which was comparable with the DCP approaches), performing the best from these groups.

7.3.1 Breakdown for sub-groups of centres

Presented here are results showing the variation of accuracy metrics for different sub-sets of centres for GCSE English language and A level biology. The intention of exploring these different breakdowns of the data are to understand differential performance of the approaches across different groups of centres. This is to understand the extent to which one approach may be achieving higher levels of predictive performance than another, by predicting more or less accurately for centres with different characteristics.[57]

For all the plots in this section which show the breakdown of predictive performance for different subgroups of centres, tables are provided in Annex E indicating the number of centres in each category.

[57] It should be noted that approaches 4 and 5 are omitted from subsequent analysis due to a lack of confidence in the accuracy of the data provided at the time of the analysis. Given the performance of these approaches shown in Figures 7.3 and 7.4, investigating this issue was been deprioritised as it is highly unlikely that either approach would be identified for implementation.

Centre size

Figures 7.5 and 7.6 show the predictive accuracy of the different approaches for groups of centres based on entry size for GCSE English language and A level biology, respectively.

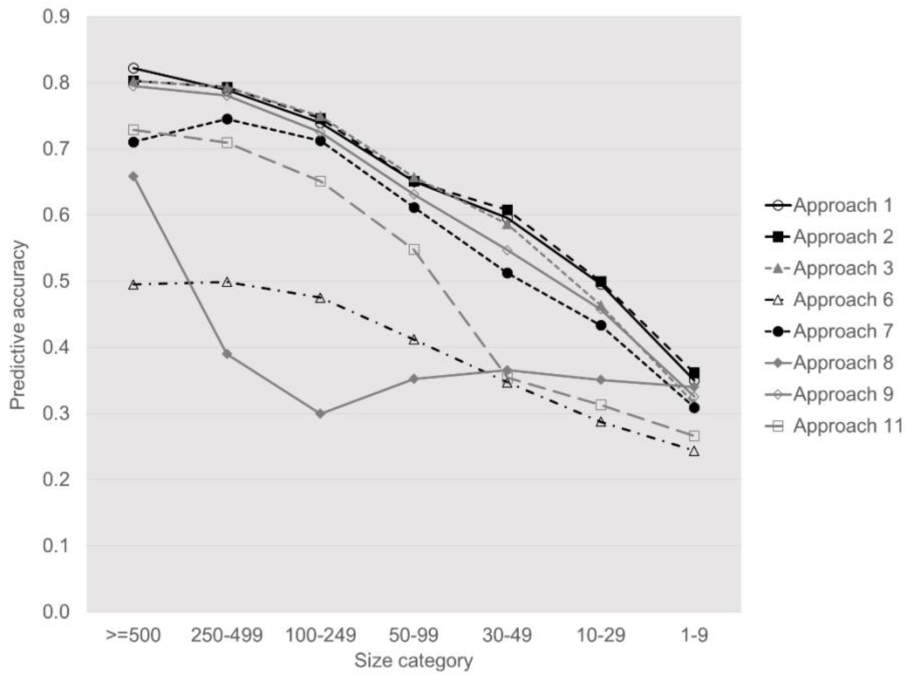


Figure 7.5 Predictive accuracy for GCSE English language by centre entry size

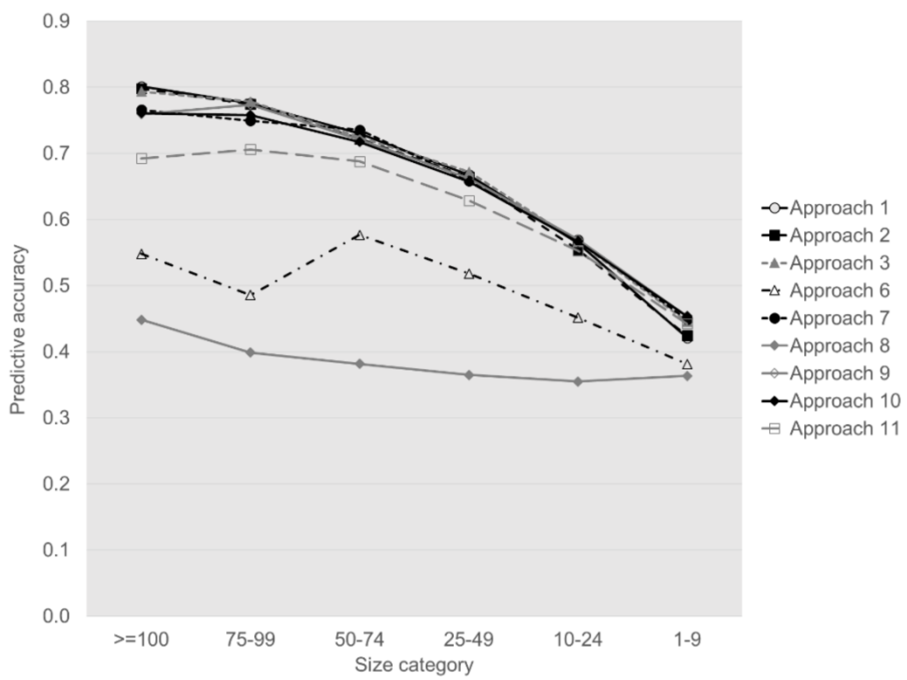


Figure 7.6 Predictive accuracy for A level biology by centre entry size

As expected, as the size of entry from the centre in the subject reduces, on the whole, so does the predictive accuracy. The main point to note from these figures is that the DCP approaches (Approaches 1 to 3) appear to consistently either outperform the other approaches or are among the strongest performing approaches across the entry sizes. Further consideration of how centres with small number of students entering for a qualification are handled in live operation is provided in Section 8.4.

Breakdown by centre outcomes

This section considers the impact of previous centre performance on predictive accuracy. This previous performance is characterised here by the centres' cumulative percentage outcomes in the subject in summer 2018.

Figures 7.7 and 7.8 show the predictive accuracy of the different approaches for GCSE English language at grades 7 and 4. Equivalent plots are provided in Figures 7.9 and 7.10 for A level biology at grades A and C.

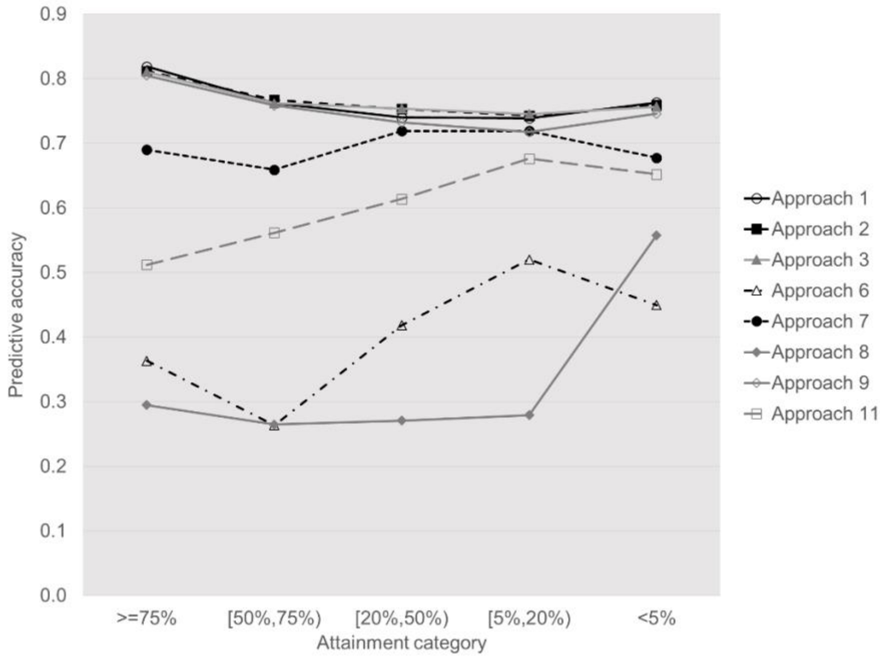


Figure 7.7 Predictive accuracy for GCSE English language by cumulative percentage outcomes at grade 7

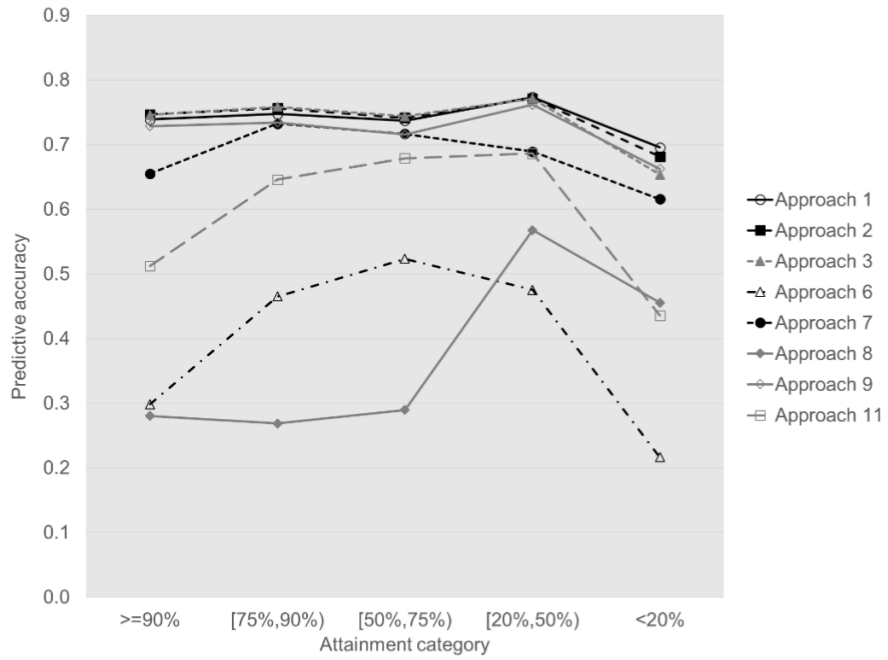


Figure 7.8 Predictive accuracy for GCSE English language by cumulative percentage outcomes at grade 4

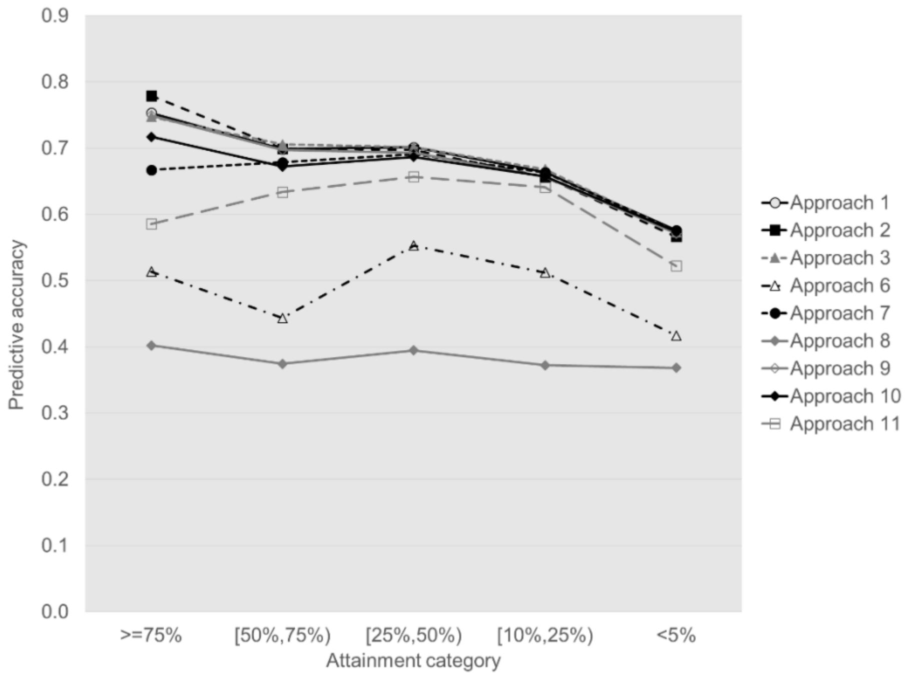


Figure 7.9 Predictive accuracy for A level biology by cumulative percentage outcomes at grade A

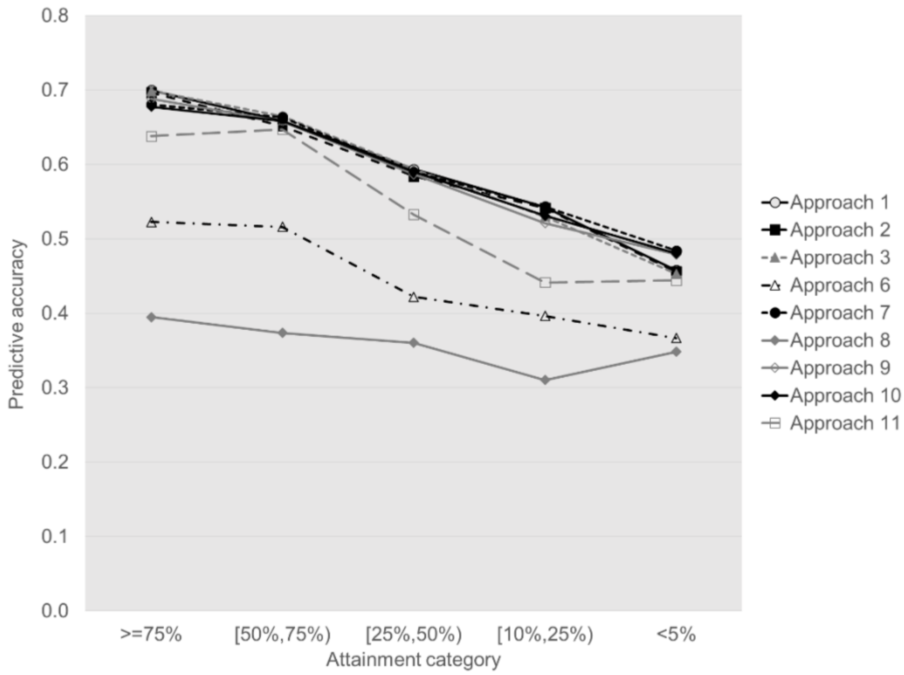


Figure 7.10 Predictive accuracy for A level biology by cumulative percentage outcomes at grade C

For GCSE English language, for the better performing approaches, the figures show a reasonably constant performance across the previous centre attainment categories. This is less the case for A level biology where there is a reduction in predictive accuracy as the previous attainment reduces. It should be remembered, however, that this is not an indicator of advantage or disadvantage to these groups given the unbiased nature of the estimates.

It can again be seen from these figures that the DCP approaches (Approaches 1 to 3) not only perform the best across the range of centre outcomes, by they are also the most consistent across the categories, particularly for GCSE English language.

Match rate of the target age group

The next set of analyses consider the performance of the approaches with varying levels of match rate to prior attainment for students who are at the target age for the qualifications: 16-year-old students matched to KS2 scores for GCSE and 18-year-old students matched to GCSE performance for A level. These are shown in Figures 7.11 and 7.12.

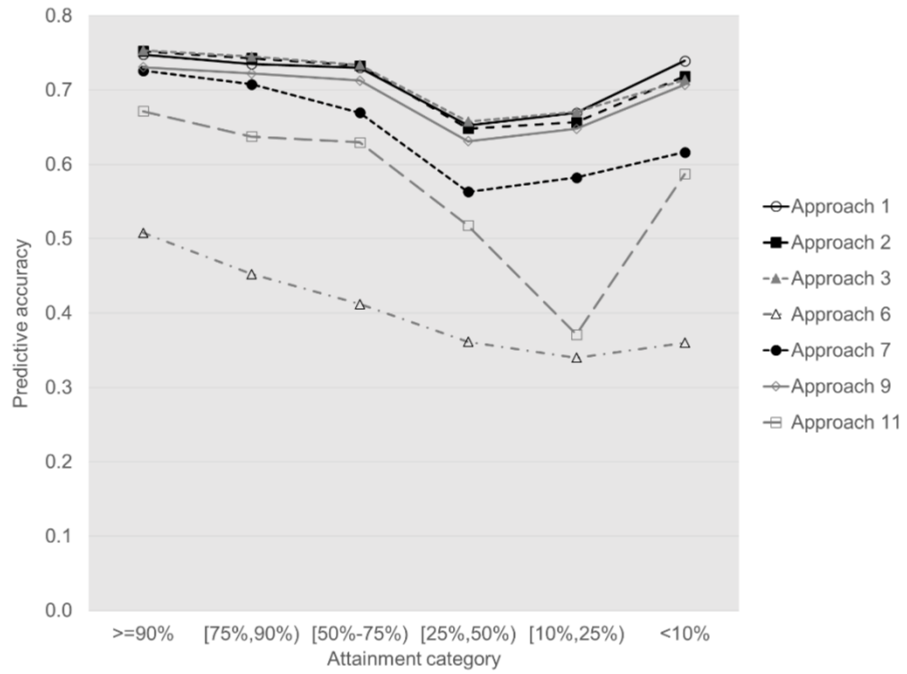


Figure 7.11 Predictive accuracy for GCSE English language by prior attainment match rate for 16-year-old students.

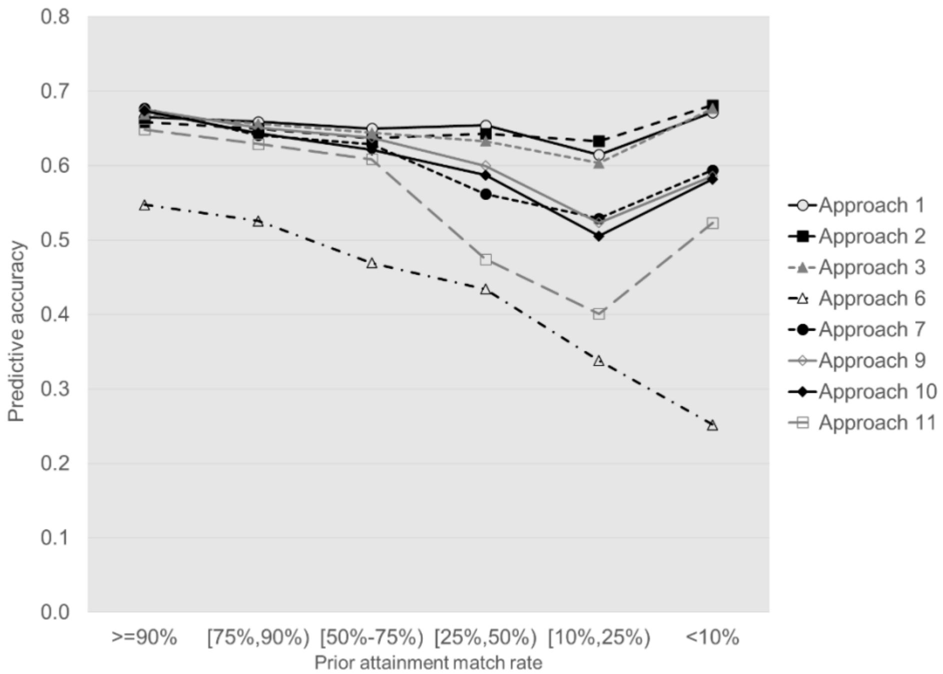


Figure 7.12 Predictive accuracy for A level biology by prior attainment match rate for 18-year-old students.

The main point of note from these figures is the tendency for the predictive accuracy to reduce slightly (for all but one of the approaches) for categories [25%, 50%) and [10%, 25%). However, once again, the DCP approaches performed consistently well across the categories.

7.4 Practical considerations and selection of the approach

Given the criticality of ensuring that students are awarded grades on time and the importance of ensuring correct and consistent application of the approach by exam boards (see the aims stated in Section 4.1), the practicalities of taking different approaches are also important to consider. The need to deliver the approach for up to 157 qualifications in parallel (see Annex A) is an important consideration.

From a delivery perspective, there is a key difference between the regression approaches and the DCP approach. The regression approaches rely on the fitting of a conventional statistical model. This contrasts with the DCP approach which uses a technique based on outcome/prediction matrices (introduced in Section 5 in the context of cohort level standards and described in detail in Section 8). While these matrices represent a ‘model’ of the value-added relationship between students’ prior attainment and the grades they are expected to achieve, they are constructed by following a mathematical procedure, rather than conventional model fitting. As highlighted above, the implementation of some of the approaches that rely on fitting models is not possible due to a failure of those models to converge. Where these models did converge, there was, however, a clear difference in computational demands of the DCP approach compared to the alternatives.

An additional factor to consider was the need for individual exam boards to implement the process in a common way. While the challenge of ensuring that exam boards share coefficients/estimates resulting from the fitting of different models is clearly not insurmountable, it does add a degree of operational complexity that is absent from the procedural solution provided by the DCP approach.

From all practical perspectives, the DCP approach was favourable from those tested. This is helpfully supported by the technical evidence outlined in the previous sections that demonstrated both high predictive performance of the approach and consistently higher performance across compared with the alternative approaches.

When selecting the approach to be applied this summer, it was important to reflect on the aims of the standardisation process discussed in Section 4.1. These are restated below for convenience:

i. to provide students with the grades that they would most likely have achieved had they been able to complete their assessments in summer 2020
ii. to apply a common standardisation approach, within and across subjects, for as many students as possible
iii. to protect, so far as is possible, all students from being systematically advantaged or disadvantaged, notwithstanding their socio-economic background or whether they have a protected characteristic
iv. to be deliverable by exam boards in a consistent and timely way that they can quality assure and can be overseen effectively by Ofqual
v. to use a method that is transparent and easy to explain, wherever possible, to encourage engagement and build confidence

Taking into account the predictive performance of the models, the practicalities of implementation and the more intuitive nature of the approach it was decided that the DCP (Approach 1) and DCP modified (Approach 3) methods would be taken forward into the centre-level equalities analysis. This analysis is presented below and seeks to identify whether there are any adverse consequences on groups with different demographic and socio-economic characteristics of applying either approach.

7.5 Centre-level equalities analysis

To inform the final decision over the approach to be taken, it was necessary to evaluate the equalities impact of these approaches. This was achieved by considering how the approaches perform across centres with differing proportions of students with different demographic and socio-economic characteristics. To ensure fairness, it is vital that particular groups of students are not disadvantaged by the standardisation approach.

7.5.1 Existing evidence on educational achievement

The presence of differences in educational achievement in England has been a matter of longstanding concern. As well documented by Strand[58], among others, the most prevailing of these inequalities are those associated with gender, ethnicity and socio-economic status. The latter was found to be the dimension along which the attainment gap is the largest. The research also highlights the presence of substantial interactions among these factors, particularly between (i) ethnicity and socio-economic status and (ii) between ethnicity and gender.

More recent data published by the Department for Education (DfE)[59] shows that by using free school meal (FSM) eligibility as an indicator of socio-economic deprivation, White students from less affluent families are among the lowest achievers in GCSE maths and English. The DfE report also shows the role of the interaction between ethnicity and gender on the attainment of students.

These differences in outcomes likely reflect differences in societal and educational opportunities. However, it is important that any approach adopted in summer 2020 to standardise CAGs does not exacerbate any of these attainment gaps by introducing assessment bias. In other words, the standardisation process is not intended to reduce existing inequalities but it should not widen (or for that matter narrow) existing attainment gaps by introducing assessment bias into the process beyond what may be considered natural variation.

7.5.2 The standardisation approach

The standardisation of CAGs will be performed at centre level and will based on two main factors:

(i) the prior attainment of students in each centre

(ii) the centre performance in recent years

[58] Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: intersectionality and ‘getting it’ for the white working class. *Research Papers in Education*, 29:2, 131-171.

[59] Department for Education (2020*). Key stage 4 performance, 2019 (revised)*. London, UK: Department for Education.

The effect of any demographic and socio-economic characteristics will be reflected in outcomes so using the previous performance at centre level as a key element of the standardisation process may be enough to account, although indirectly, for the demographic and socio-economic centre level composition. However, it may be that previous centre performance and the prior attainment of the current cohort of students are not enough to control for the attainment gaps, especially if the proportions of such students within centres has changed over time.

To understand if accuracy of the standardisation approach is affected by students' demographic/socio-economic characteristics, it is necessary to check empirically whether any demographic and socio-economic group is disadvantaged by the standardisation process.

To gain a better understanding of the equalities impact of the standardisation approach, there are three research questions that need to be addressed as outlined in Table 7.2.

Table 7.2 Centre-level analysis research questions

Question:	Addressed by:
Is there any effect of centre-level demographic and socio-economic composition on the ***accuracy of predicted grades***?	Evaluating the accuracy of predictions as tested with respect to 2019 outcomes
Is there any impact of centre-level demographic and socio-economic composition on ***variability in year-on-year centre-level outcomes***?	Comparing: 1. Fluctuations over time due to the standardisation approach with 2. Year-on-year fluctuations in outcomes that normally occur
Which one of the two approaches is showing a ***better performance*** in terms of bias in predictions for different demographic and socio-economic groups?	Comparing analysis results obtained for the DCP approach and the DCP Modified approach

7.5.3 Methodology

The data combines the centre-level outcomes and a set of variables describing the demographic and socio-economic composition of students taking each subject in each centre. Predicted outcomes have been retrieved from the testing exercise described in Section 7.3. The demographic and socio-economic composition of each centre was retrieved from the National Pupil Database (NPD).

In line with the research questions above, there are three main dependant variables of interest. These are the dimensions that may be affected by the centre-level demographic/socio-economic composition and/or by how the centre demographic/socio-economic make up has changed over time. These are the differences between:

A. Predicted outcomes – Actual outcomes in 2019

B.1. Predicted outcomes in 2019 – Actual outcomes in 2018

B.2. Actual outcomes in 2019 – Actual outcome in 2018

A. concerns the accuracy of predictions and allows evaluation of whether any demographic and socio-economic group of students are disadvantaged by the use of the standardisation approach.

B. relates to year-on-year changes in outcomes, a comparison of the results of the statistical standardisation process (B.1) and actual outcomes (B.2) obtained in 2019 with respect to 2018 actual outcomes. If changes in the attainment gap occurred between 2018 and 2019 it is unrealistic to expect that the standardisation approach can predict these changes.

Centre-level outcomes were evaluated with respect to:

i. For GCSE, at grade 7 (or above) and grade 4 (or above)

ii. For A level, at grade A (or above) and grade C (or above)

Demographic and socio-economic centre compositions were defined on the basis of:

- Gender: proportion of female students;
- SEN status: proportion of students with Special Education Needs;
- EAL status: proportion of students with English as Additional Language;
- FSM eligibility: proportion of students eligible for Free School Meals among those who claimed it;
- IDACI score: proportion of students in the first tertile (33% highest deprived students) and proportion of students in the first and second tertile (66% highest deprived students);
- Major ethnicity group: proportion of white and non-white students;
- Minor ethnicity grouping: a breakdown of non-white students into a more refined categorisation (considered only for large-entry subjects);

On the basis of recent literature on attainment gaps, a further breakdown was also considered based on two-way interactions between the demographic and socio-economic variables listed above:

- Major ethnicity group and gender;
- Major ethnicity group and FSM.

Although other interactions, including three way interactions (eg ethnicity, gender and FSM) could be included in the analysis in principle, this would entail deriving proportions based on very small numbers which may be unreliable. For this reason these smaller groups were not considered.

The analysis was conducted for the subjects listed in Table 7.3 below:

Table 7.3. Subjects included in centre-level equalities analysis

GCSE	A level
English language	Biology
History	French
Geography	Religious studies
Maths	
Music	

Detailed findings are reported for GCSE English language and a summary of findings for the other subjects are also reported. All results are reported in Annex F (which includes the results related to research question A on the impact on accuracy of predictions) and Annex G (which includes the results related to research question B on the impact on variability in year-on-year centre-level outcomes).

The intention was to explore the relationship between the demographic/socio-economic characteristics of students and the accuracy of the standardisation approaches, to compare differences between predicted and actual grades to the fluctuations that are likely to occur in any case. Given that the standardisation of CAGs was designed to be applied at subject/centre-level, the analysis was performed at centre-level and separately for each subject.

In addition to presenting a set of descriptive statistics, the analysis includes a regression analysis. This allowed exploration of the effect of each demographic/socio-economic characteristic, when all the other factors were held fixed. The regression specification took the form:

$$Y_i = \alpha + \beta X_i + \gamma Z_i + e_i$$

where:

- Y_{ij} was the dependant variable for centre i summarising changes in outcomes. As described above this can take three different forms (as defined in A., B.1., B.2.);
- X_i was a set of independent variables summarising the demographic and socio-economic centre-level composition (described above). It should be noted that when the dependant variable was the fluctuation over time in outcomes (research questions B.1 and B.2), these independent variables were also measured as the difference between the centre composition in 2019 and 2018. These variables were centred around the mean.
- Z_j was a set of additional centre-level variables controlled for in the regression analysis (ie, size and average prior attainment);
- α, β, γ were the regression coefficients;
- e_i was the error term - assumed to be independent and identically distributed.

Within the above regression specification, it is β that yields the relationship between each demographic/socio-economic characteristic and the dependant variables. The estimate of this parameter was mostly of interest.

7.5.4 Results

GCSE English language

Starting with research question A, Figure 7.13 – panel *a.* considers the distribution of the prediction accuracy for GCSE English language, measured as the difference between predicted and actual outcomes in 2019 at grade 7 (or above) and grade 4 (or above), separately for the DCP and the DCP Modified approach. The prediction accuracy was centred around zero and followed a symmetric distribution, indicating that there was no systematic error in predictions. Figure 7.13 – panel *b.* displays the distribution of the demographic and socio-economic characteristics considered as part of this analysis. The blue vertical line indicates the mean of the distribution.

Figure 7.14 shows four sets of scatter plots, each representing the correlation between a demographic and socio-economic characteristic and the accuracy of predictions, at grade 7 (or above) and grade 4 (or above), separately for the DCP and DCP modified approach. For each graph the Pearson correlation coefficient is also reported, along with a smoothed curve[60] showing the relationship between each pair of variables.

The correlation coefficients were mostly very low, and lower for the DCP than for the DCP Modified approach. Only for a few demographic and socio-economic characteristics, were the correlation coefficients as high as 0.15-0.16 (for the DCP Modified approach at grade 4). This indicated a weak link between accuracy and gender, SEN status, FSM eligibility, the interaction between gender and ethnicity, and the interaction between ethnicity and FSM eligibility.

A more in-depth analysis of the impact of demographic and socio-economic centre-level composition was provided by the regression analysis in Table 7.4 which shows the impact on accuracy at grade 7 (or above) and grade 4 (or above) of a one-percentage point change in each of the variables considered. For example, with reference to the DCP approach, a 1 percentage point (pp) increase in the share of female students can lead, on average, to an underestimation of 0.02 pp in the proportion of grade 7 (or above) and an underestimation of 0.08 pp at grade 4 (or above). These are very small effects. For the DCP Modified approach the corresponding figures are larger, 0.03 and 0.14, respectively.

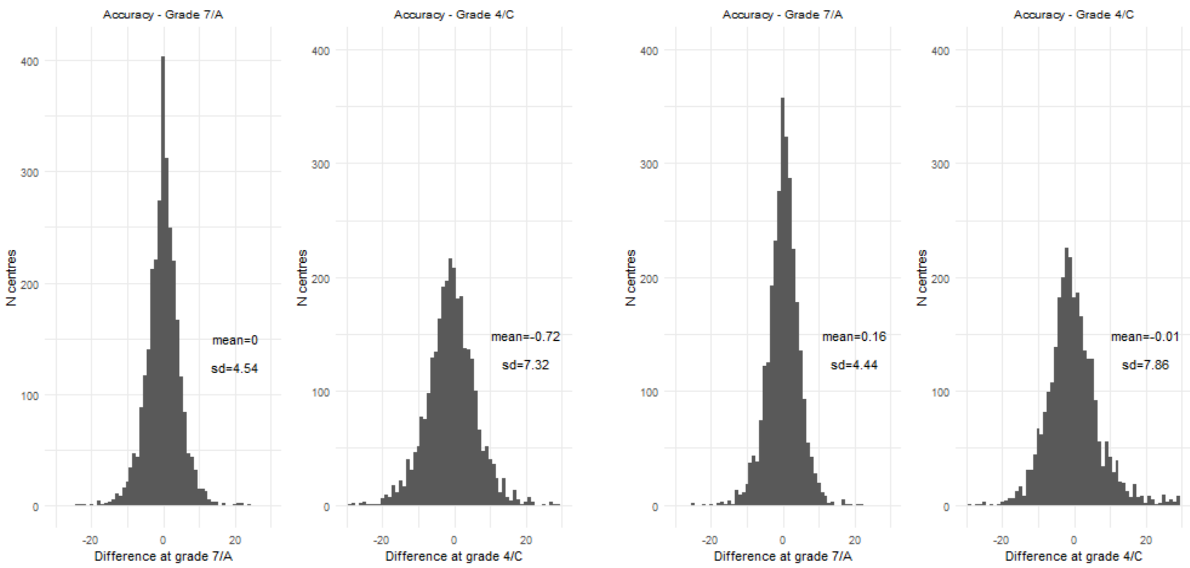
Although statistically significant, changes in the share of students with SEN status and different socio-economic deprivation (IDACI score) also had very small effects on the prediction accuracy of the DCP approach. There was no statistically significant effect of changes in ethnicity or FSMs on the accuracy of the DCP approach.

60 This is a LOESS (Locally Estimated Scatterplot Smoothing), a non-parametric regression which represents the relationship between two variables without imposing a structural form to the data.

Figure 7.13. Univariate distribution of key variables – GCSE English language

a. Distribution of accuracy at centre level

DCP ***DCP Modified***



b. Distribution of demographic and socio-economic centre composition

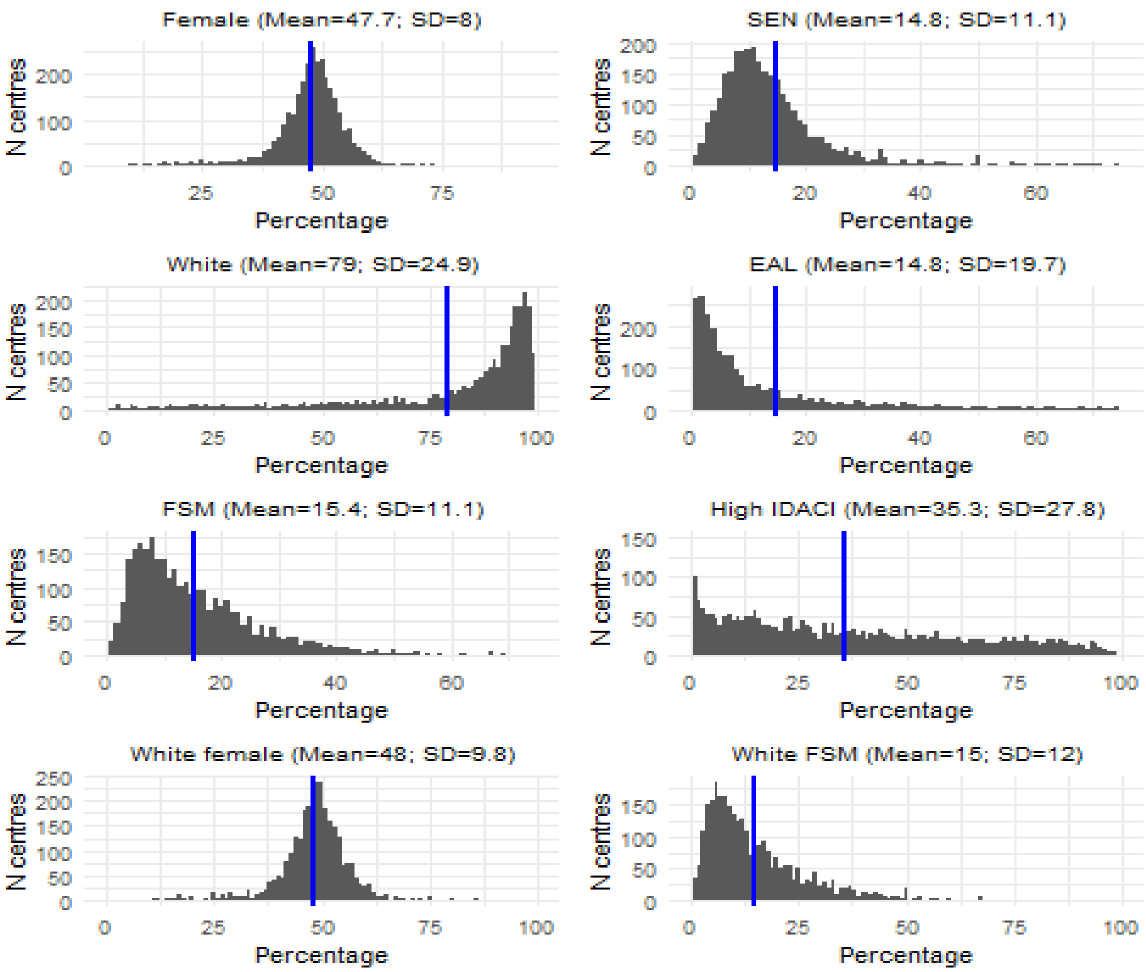


Figure 7.14. Correlation between demographic/socio-economic factors and accuracy of predictions – GCSE English language

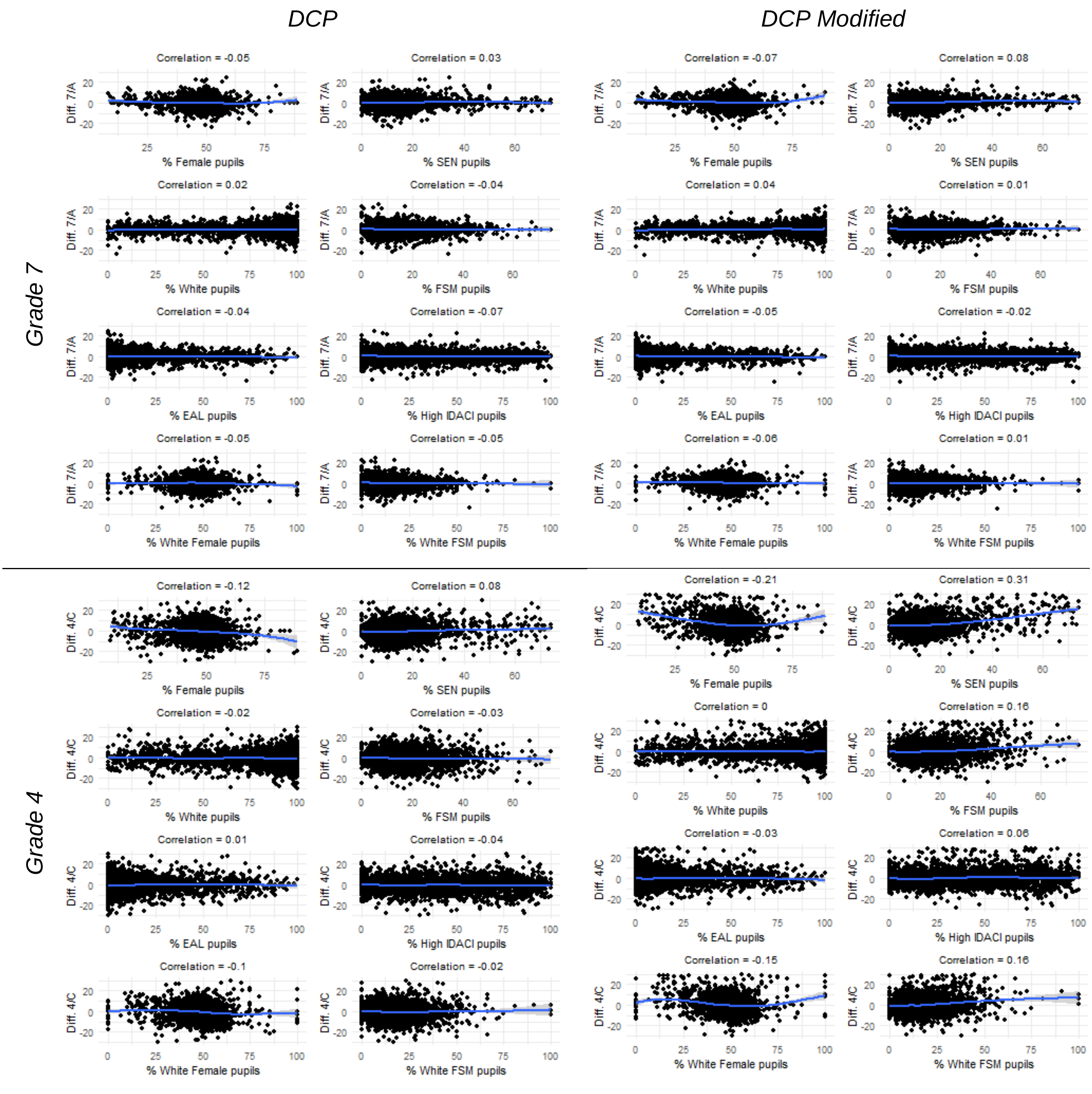


Table 7.4. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – GCSE English language

	DCP		DCP Modified	
	Acc. 7/A+	Acc. 4/C+	Acc. 7/A+	Acc. 4/C+
(Intercept)	0.07	-0.77 *	0.49 **	1.66 ***
	(0.19)	(0.30)	(0.18)	(0.31)
Perc.Female	-0.02	-0.08 **	-0.03	-0.14 ***
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.SEN	0.02	0.05 ***	0.02 **	0.15 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.WHITE	-0.01	-0.02	-0.01	-0.04 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.FSM	0.01	-0.06	0.01	-0.01
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.EAL	-0.01	-0.00	-0.02 *	-0.05 ***
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.high	-0.01 *	-0.02 *	-0.00	-0.02 **
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.med	0.00	-0.00	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
WHITE_Female	-0.00	-0.01	-0.01	0.01
	(0.01)	(0.02)	(0.01)	(0.02)
WHITE_FSM	-0.02	0.02	-0.01	0.05
	(0.02)	(0.03)	(0.02)	(0.03)
mean.KS2	-0.24	-1.87 *	-0.56	-3.49 ***
	(0.55)	(0.88)	(0.53)	(0.88)
Ncands_2019	-0.00	0.00	-0.00 *	-0.01 ***
	(0.00)	(0.00)	(0.00)	(0.00)

*** p < 0.001; ** p < 0.01; * p < 0.05.

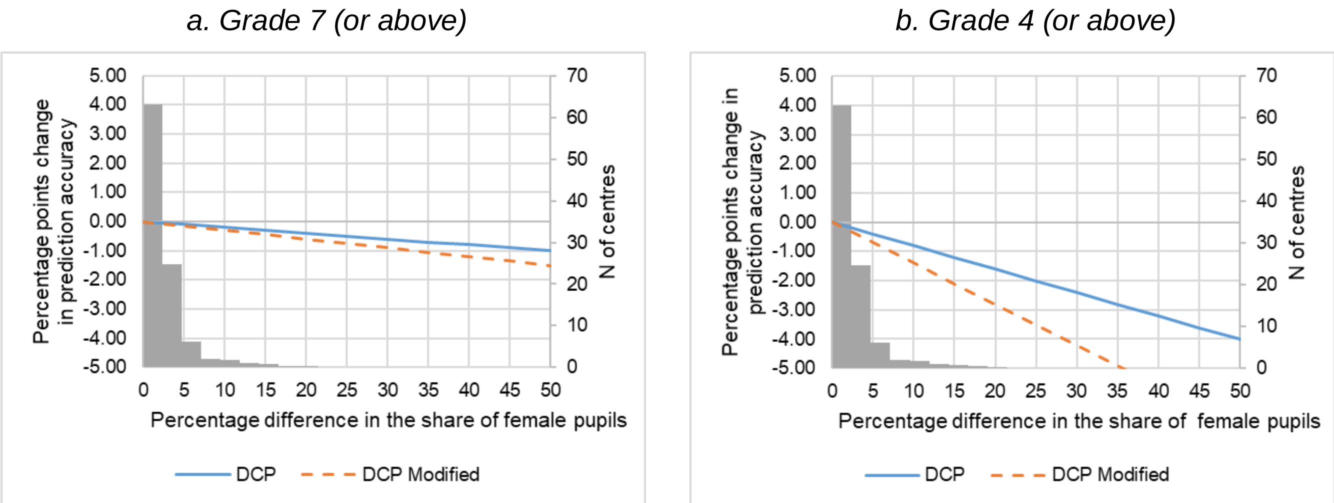
Focusing on gender, the variable from Table 7.4 that seems to have the largest impact on prediction accuracy[61], it is possible to evaluate the size of its impact in different scenarios. Figure 7.15 reports the impact on accuracy by different variations from the mean in the share of female students (see the straight lines, blue solid for

[61] This is excluding prior attainment (mean KS2), that is having a larger impact. However, prior attainment is not one of the main independent variables, only one control variable added to the regression specification to account for what is known to be the strongest predictor of attainment.

the DCP approach and orange dotted for the DCP modified approach). For both the DCP and the DCP modified approach, the impact was negative. Once all the other factors are held fixed, for different changes in the share of female students, the DCP approach was associated with smaller changes in outcomes than the DCP Modified approach.

In Figure 7.15 the distribution of centres by share of female students is also reported in the form of vertical bars. This indicates that there were not many centres with a share of female students very different from the mean (47.7, indicated as 0 in the graph). Focussing on the DCP approach results, this means that, at grade 7 (or above), a different gender composition is unlikely to have any substantial impact on the accuracy of results. At grade 4 (or above), given that the impact was slightly larger, only centres with 12.5% more female students than the average centre would potentially be underpredicted by the standardisation approach by more than 1 pp. This would amount to around 8% of centres.

Figure 7.15. The marginal effect on accuracy of a variation from the mean in the share of female students



Moving to research questions B.1 and B.2, Figure 7.16 displays the year-on-year changes in outcomes at centre level. Centre-level variability is reported separately for the DCP approach (panel *a.*), the DCP modified approach (panel *b.*), and the actual outcomes (panel *c.*). The latter refers to research question B.2 as it reports the actual centre-level variability in outcomes between 2018 and 2019.

From the graphs reported in Figure 7.16 it is clear that the actual variability in outcomes between 2018 and 2019 (panel *c.*) was greater than the variability in outcomes resulting from the use of the predicted results for 2019, both at grade 7 and grade 4. This is true for both the DCP and the DCP modified approach, although for the DCP modified approach the heavier tails of the distribution suggest a larger standard deviation and therefore some inaccuracy.

Figure 7.17 shows the correlation between each demographic and socio-economic characteristic and the year-on-year variability of results at grade 7 (or above) and grade 4 (or above), for the DCP and DCP Modified approach, to be compared with the variability that actually occurred between 2018 and 2019. The correlations were very low. Importantly, the correlations tended to be higher when the actual variability in results is considered, suggesting that the impact on year-on-year centre variability of results may be smaller in 2020 (when standardisation will occur) than in a normal year. This result suggests that the standardisation approach will not widen any attainment gap.

The output of the regression analysis for centre-level variability is presented in Table 7.5. This confirmed the main points highlighted by the bi-variate correlations shown in Figure 7.17. Overall, there was a very small impact of changes in demographic and socio-economic composition on over time centre variability. In particular, the impact of each of these characteristics was larger when actual outcomes were considered compared with when predictions from either the DCP or DCP modified approach were used. Between the two standardisation approaches, DCP was less affected by changes in centre composition.

In sum, share of female students had the largest impact on year-on-year variability. For the DCP approach, the estimated impact of one-percentage point increase in the share of female students was 0.02 and 0.06 pp, at grade 7 and grade 4 respectively. However, centres would need to be large to experience any change in outcomes as a result of a change in the share of female students.

Figure 7.16. Year-on-year variability in outcomes – GCSE English language

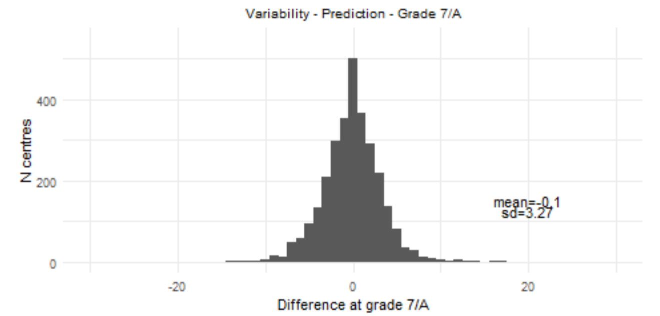
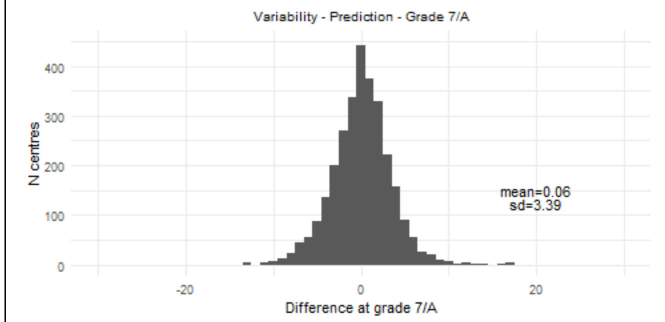
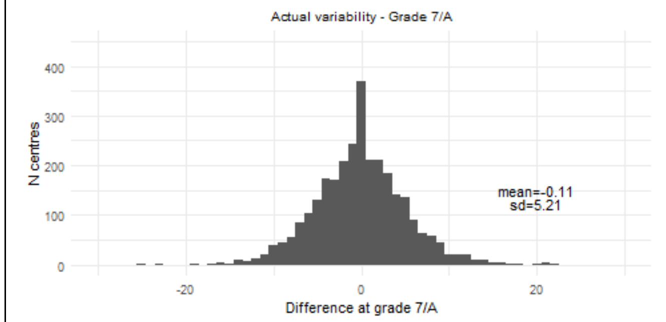
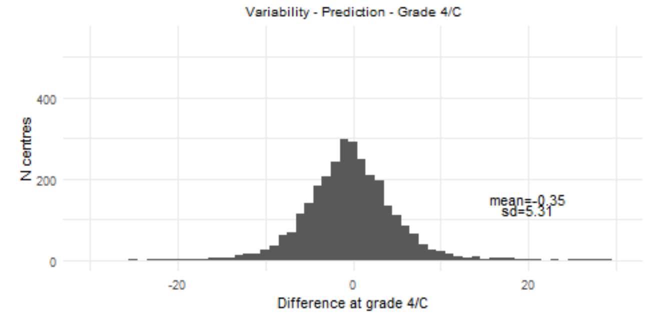
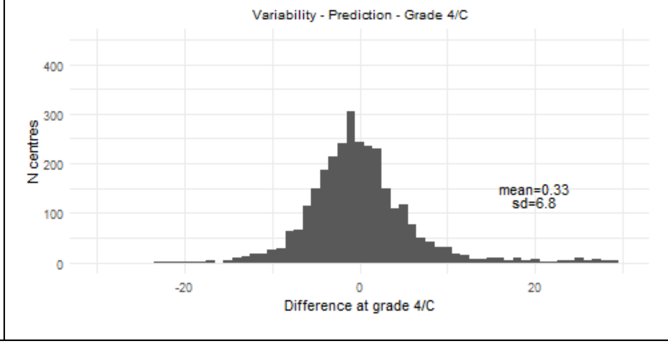
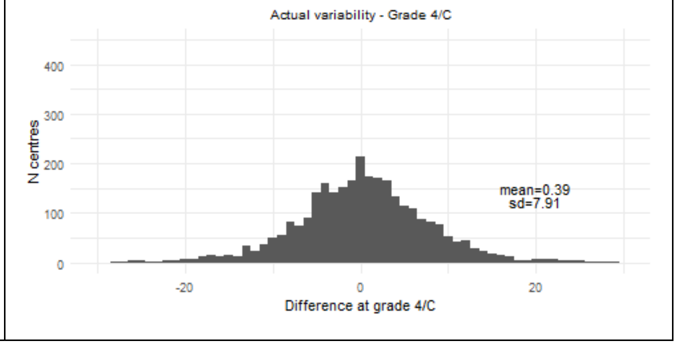
	a. DCP	*b. DCP Modified*	*c. Actual variability (B.2)*
Grade 7/A			
Grade 4/C			

Figure 7.17. Correlation between demographic/socio-economic factors and accuracy of predictions – GCSE English language

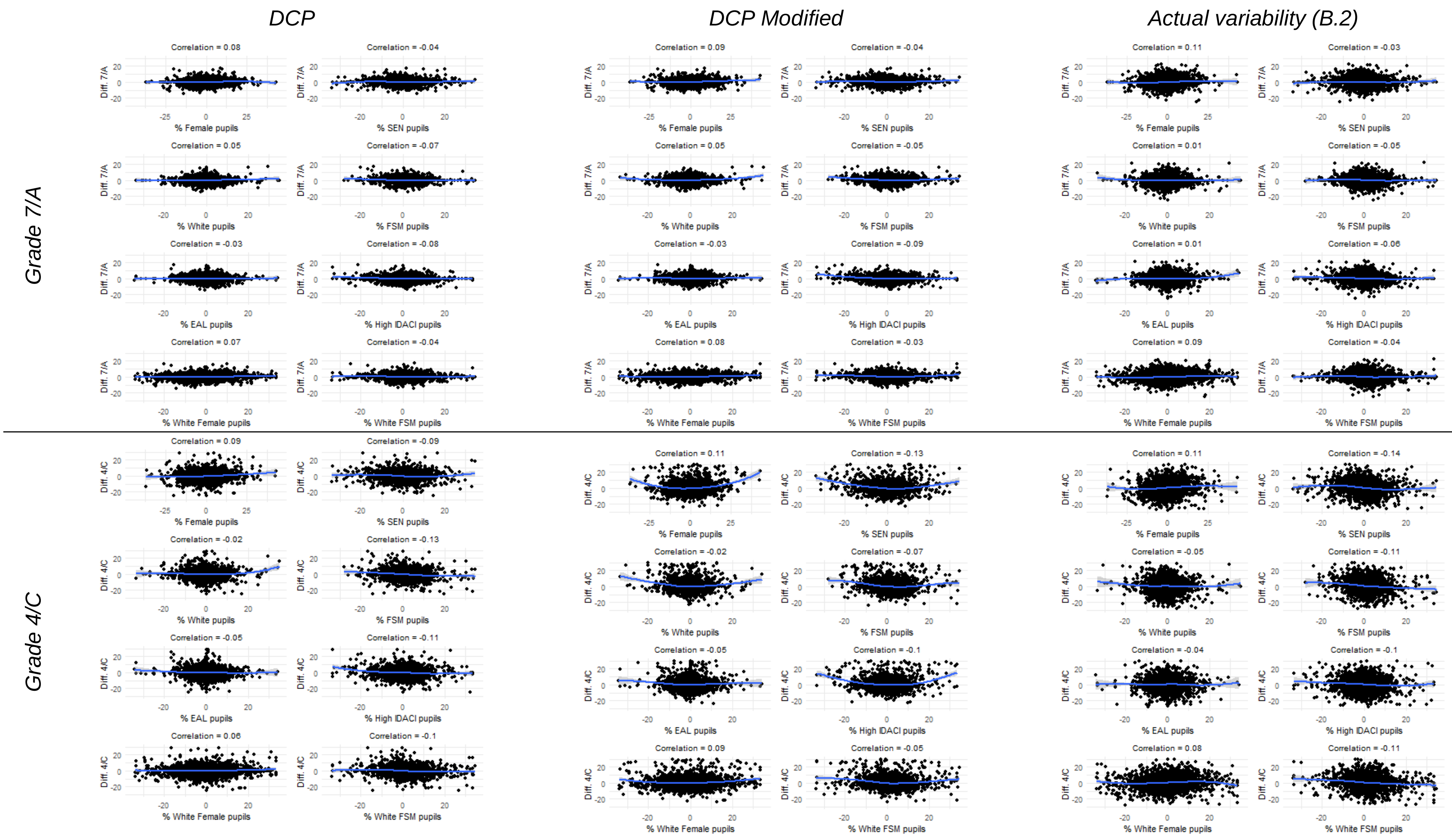


Table 7.5 Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – GCSE English language

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	Var. 4/C+	Var. 7/A+	Var. 4/C+	Act. 7/A+	Act. 4/C+
(Intercept)	-0.30 *	-0.35	0.28 *	3.24 ***	-0.38	0.45
	(0.13)	(0.19)	(0.13)	(0.25)	(0.21)	(0.31)
Perc.Female	0.02	0.06 ***	0.02 *	0.08 ***	0.05 **	0.12 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Perc.SEN	0.00	-0.01	0.00	-0.07 ***	0.00	-0.09 ***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Perc.WHITE	0.02 *	-0.03	0.02 *	-0.04 *	0.01	-0.08 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Perc.FSM	0.00	-0.02	0.02	-0.00	0.02	-0.01
	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)
Perc.EAL	0.00	-0.02	0.00	-0.04	0.03	-0.04
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
IDACI.high	-0.03 **	-0.03 *	-0.04 ***	-0.03	-0.07 ***	-0.10 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
IDACI.med	-0.02	0.03	-0.01	0.06 **	-0.05 **	-0.04
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
WHITE_Female	0.01	-0.01	0.01	0.01	0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
WHITE_FSM	0.01	0.00	0.01	0.01	-0.00	-0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
mean.ks2	8.98 ***	17.39 ***	9.60 ***	17.67 ***	9.05 ***	16.33 ***
	(0.43)	(0.67)	(0.45)	(0.87)	(0.72)	(1.06)
Ncands_2019	0.00	-0.00	-0.00	-0.02 ***	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

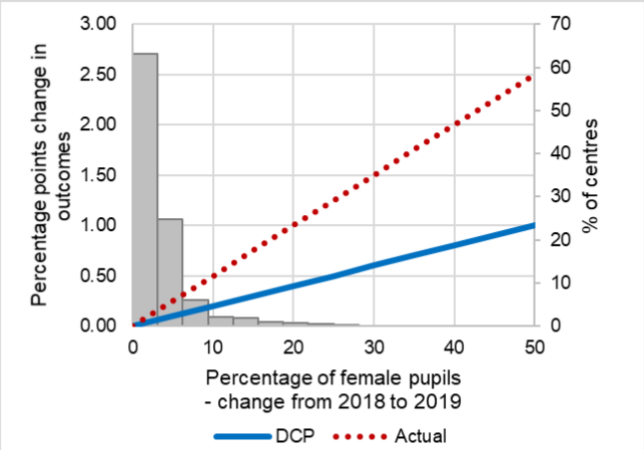
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

The impact on year-on-year variability according to changes in the share of female students, when other factors are controlled for, is displayed in Figure 7.18. Given that the DCP approach performed better than the DCP Modified approach, the latter is not reported.

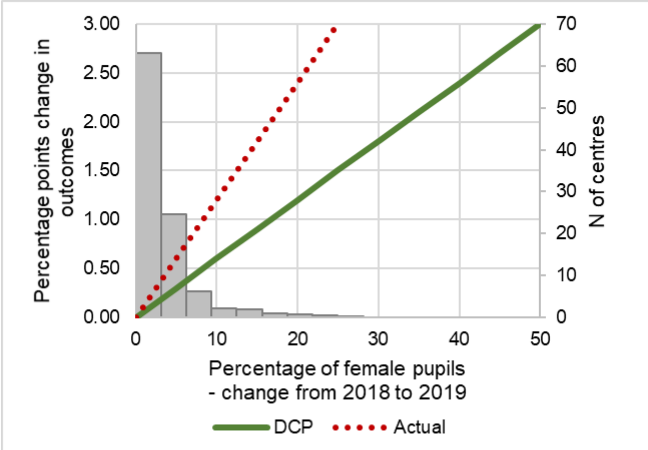
From the graphs it is clear that, for each given change in the share of female students, the year-on-year variability induced by the use of a standardisation approach was smaller than the variability in outcomes actually observed from 2018 to 2019. For example, a 20 pp increase in the share of female students was associated with an increase in the proportion of grade 7 (or above) predicted by the DCP standardisation approach of 0.4 pp, compared with an observed increase in outcomes of 1 pp. At grade 4 the impact of a 20 pp increase in the share of female students was of 1.2 pp when the outcomes predicted by the DCP approach are considered, and 2.4 pp when the actual outcomes are considered. There is, however, a very small number of centres with large changes in the share of female students from one year to the next.

Figure 7.18. The marginal effect on year-on-year variability of a change in the share of female students

a. Grade 7 (or above)



b. Grade 4 (or above)



Analysis of other subjects

Full findings for other subjects are presented in Annex F and Annex G. To summarise:

Research question A. on the effect on the accuracy of predictions

- In terms of accuracy of predictions, the impact of demographic and socio-economic characteristics was relatively small, often statistically not different from zero;
- The size of the impact was often not operationally relevant and unlikely to cause any substantial difference to the results awarded to each centre;
- The strongest effect was found for White FSM students in GCSE history[62] at grade 4, where a 10 pp increase in the share of this group of students was associated with an over prediction of 1.6 pp.

62 A level French showed some slightly larger effects in some instances, but the very small sample size makes this result not very reliable.

Research question B. on the effect on year-on-year centre level variability

- Overall, the impact of a change in demographic and socio-economic centre composition was associated with very small changes in year-on-year centre-level variability. These changes seem unlikely to cause any substantial difference to the results achieved by centres;
- The strongest effect was found for GCSE maths for students from a highly deprived background (high IDACI score). However, the impact was small and only applied to centres with dramatic increases in the share of highly deprived students which is unlikely;
- Changes in demographic and socio-economic centre composition were systematically associated with larger impacts on centre variability when the actual outcomes rather than either the DCP or the DCP modified approach were used.

Research question C. on which standardisation approach should be employed

- With respect to both accuracy of predictions and year-on-year variability, the impact of demographic and socio-economic centre composition was consistently smaller on the DCP approach than on the DCP modified approach, with one notable exception, FSM eligibility for history. This suggests that the DCP approach should be favoured.

7.5.5 Discussion

Overall, the impact of demographic and socio-economic centre composition on both accuracy of predictions and year-on-year centre variability was very small. In many cases the impact was not statistically significant. Further, the size of the effect was so small that only the results of very atypical centres, or those with large changes in demographic/socio-economic make up, would be affected. For results to be affected would also require centres to have large entries in individual subjects.

These analyses modelled a variation in outcomes (either between predicted and actual outcomes, or year-on-year changes), whereas the literature on attainment gaps focuses on the actual outcomes achieved. It is possible that some of the expected effects on attainment cancel each other out when a variation is considered. Further, the regression analysis accounts for prior attainment, which is known to be, by far, the strongest predictor of attainment and, as this analysis suggests, also of changes in attainment.

Although some of the demographic and socio-economic characteristics tended to have a slightly larger effect than others, there was not a clear pattern across subjects. This is reassuring as it suggests that there was not a specific variable that is likely to impact results of the same centres in all subjects.

Among the two approaches, the DCP approach showed consistently better results than the DCP modified approach, both in terms of accuracy and year-on-year variability in outcomes.

There are some limitations to the analysis presented. First, not all subjects have been considered. The subjects included, however, were selected to ensure breadth of coverage in terms of entry size and nature of the material taught. They are therefore considered to be representative subjects available at GCSE and A level.

Second, a regression approach was used to explore the impact of (changes in) demographic/socio-economic characteristics. The potential presence of other factors not accounted for in the analysis suggests caution in interpreting the findings. They should be interpreted as measures of association when other factors are controlled for.

Nonetheless, the evidence suggests that the standardisation approaches will not exacerbate existing inequalities in educational attainment. Of the two standardisation approaches considered, the use of the DCP approach in particular seems to minimise the equalities impact.

A discussion of the characteristics of the different approaches in relation to the aims of standardisation was provided in Section 7.4. Building into these considerations the improved performance from an equalities perspective as described here, it was decided that DCP (Approach 1) would be implemented this summer.

7.6 DCP Accuracy measures

To provide a more complete evaluation of the performance of the selected approach, the predictive accuracy for a wider range of subjects was performed. As with the analyses provided in Section 7.3, for A level, it was possible to evaluate the performance of the approaches using the 3-year span of data that was to be available operationally. This was not the case for GCSE. Only a single year of data was available for all subjects with the exception of English language, English literature and mathematics (for which it was possible to use 2 years' worth of historical data in the testing). For this reason, the accuracy metrics reported here for GCSE are also likely to be slightly lower than would be realised in 2020. Based on the limited evidence highlighted in Section 7.2, the magnitude of this effect is likely to be ~1.5% points. Also, the testing at GCSE was only possible for reform phase 1 and 2 subjects (see Annex A).

As discussed in the introduction to this section, the key metric used as a basis for comparison between approaches was the probability of an exact match between the predicted grade and the grade the student achieved. However, an additional measure that is valuable when comparing the performance of assessment systems is the level of accuracy within plus or minus one grade[63]. These measures are valuable as they reflect the unavoidable unreliability in any form of assessment. This is not something unique to the approaches it has been necessary to be apply this summer, but is relevant to any assessment system that classifies students into attainment categories (in our case defined by grades). Typically, reductions in reliability can arise from the practical limitations of domain sampling, marking variability or environmental/experiential factors. This year, the source of unreliability will be the inevitable limitations of predicting assessment outcomes in the absence of live assessment data.

Using this summary statistic, provided in Figure 7.19 is a plot of the predictive accuracy, plus or minus one grade, across A level subjects based on the analysis of 2019 data. This plot does not take into account the need to put in place special arrangements for centres with a small entry in a subject which will be necessary in

[63] Pilliner (1969) as cited in Cresswell M., (1986). Examination Grades: how many should there be? British Educational Research Journal 12, 37-54.

live operation (see Section 8.4) and will likely have a more significant effect at A level due to the likelihood of centres having a small cohort for a particular subject. To reflect a more appropriate estimate of accuracy, therefore, Figure 7.20 shows the same analysis only including centres with at least 10 students entering for the subject.

This figure shows that 51 of the 55 subjects included in the analysis achieved classification accuracies within a grade of higher than 90%.

An equivalent plot is provided in Figure 7.21 for GCSE. However, this analysis is necessarily limited in scope, as described above. For the large entry subjects reformed first, and for which the years of available historical data match that operationally available for standardisation this summer (GCSE English language, English literature and mathematics), the measures of accuracy within a grade are all ≥99%, with 12 of the 22 subjects achieving >90% accuracy within a grade.

Given the broadly comparable correlations between prior-attainment measures used at GCSE and A level, this reduction at GCSE is likely to be predominantly due to the increased number of grades available for the qualification, which is known to increase misclassification rates[64].

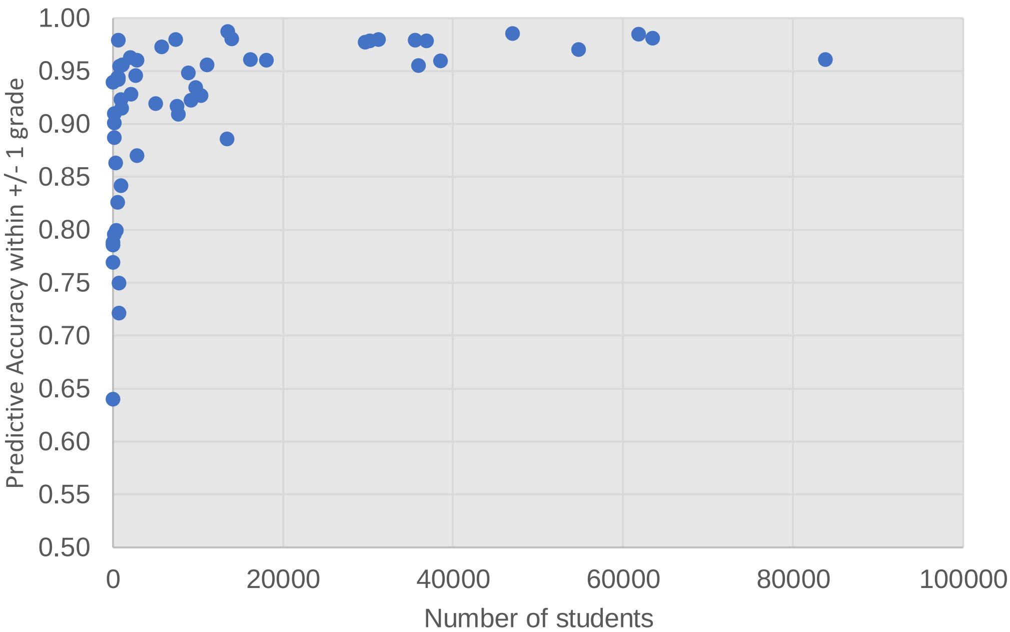


Figure 7.19 Predictive accuracy +/- 1 grade across A level subjects

[64] Cresswell M., (1986). Examination Grades: how many should there be? British Educational Research Journal 12, 37-54.

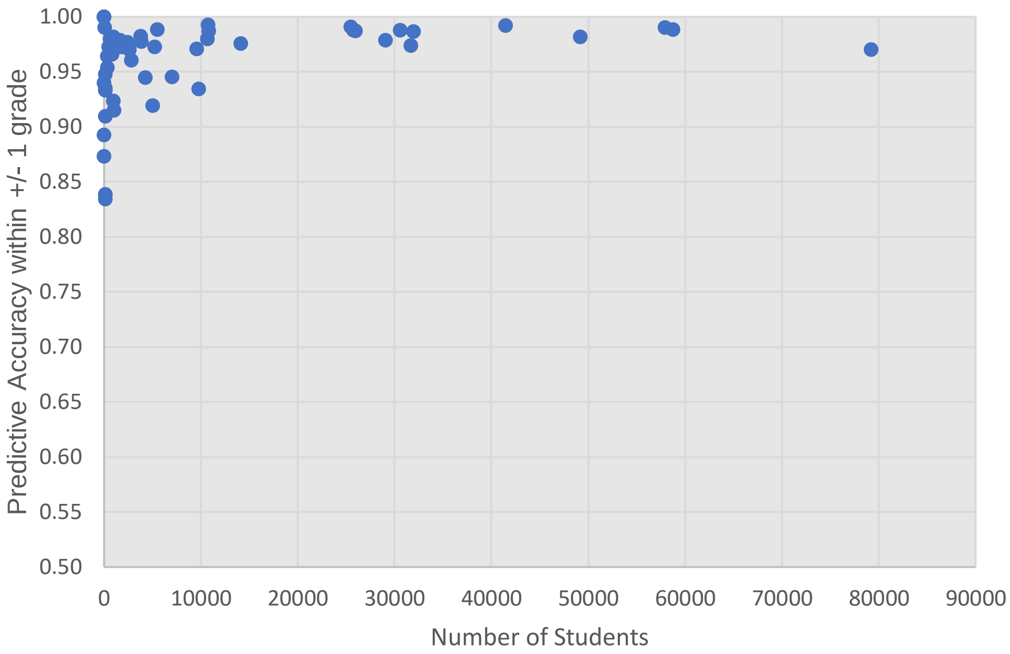


Figure 7.20 Predictive accuracy +/- 1 grade across A level subjects excluding centres with fewer than 10 students

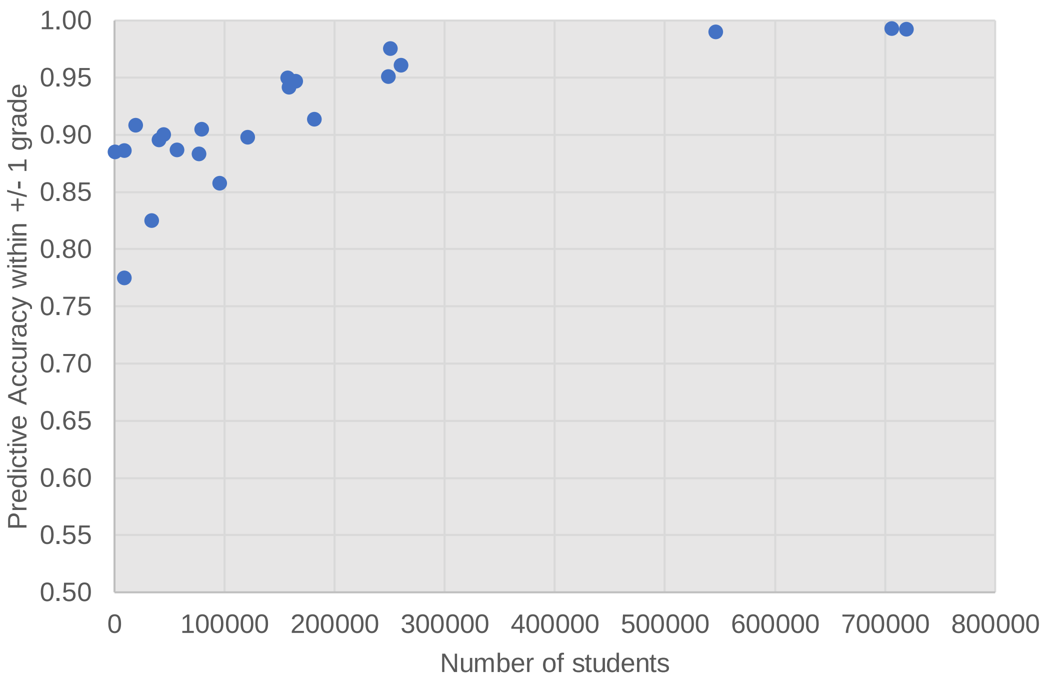


Figure 7.21 Predictive accuracy +/- 1 grade across GCSE subjects

Shown in Figures 7.22, 7.23 and 7.24 are the predictive accuracy measures for: all A level subjects (all centres); all A level subjects (centres with fewer than 10 students removed); and for the 22 GCSE subjects for which some post-reform historical data were available. These are each plotted against the number of students in the data set for each subject. The data corresponding to these plots are provided in Annex E.

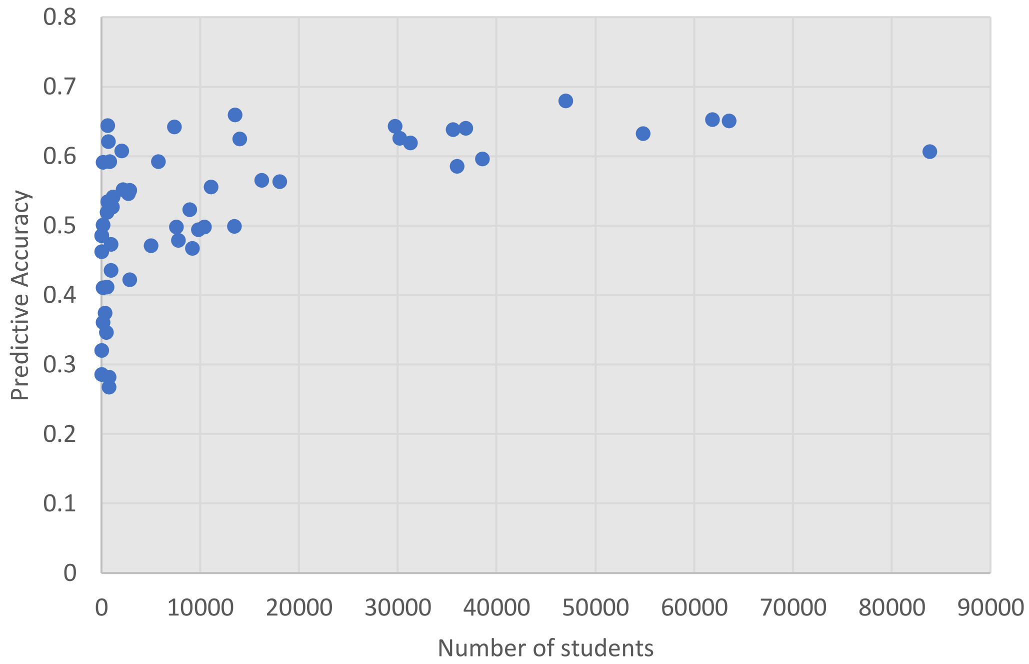


Figure 7.22 Plot of predictive accuracy by entry size for a range of A level subjects

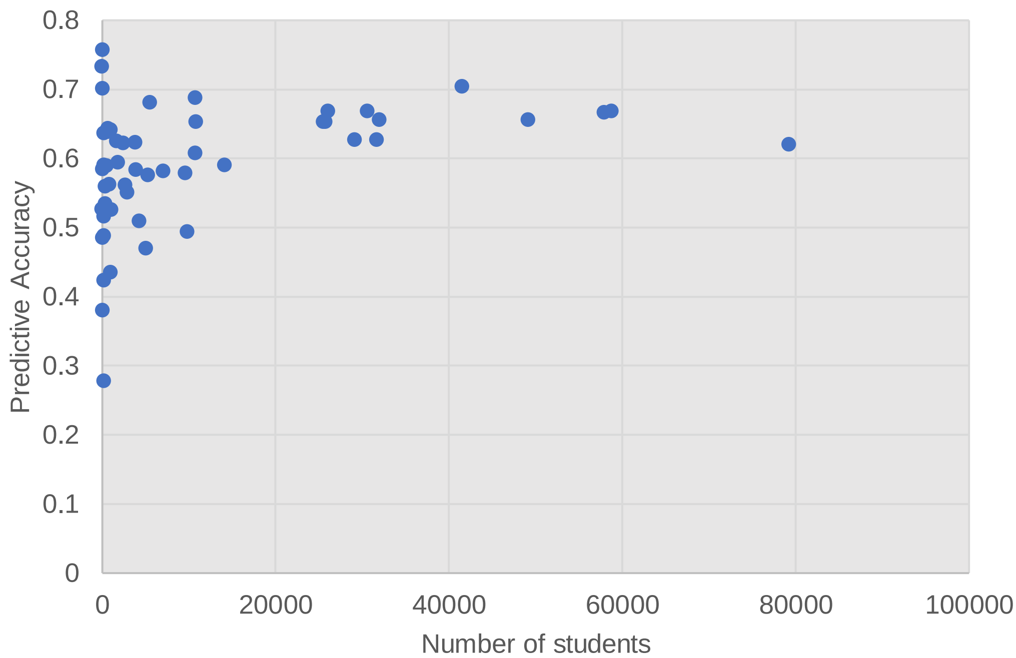


Figure 7.23 Plot of predictive accuracy by entry size for a range of A level subjects with centres with fewer than 10 entries removed from the analysis

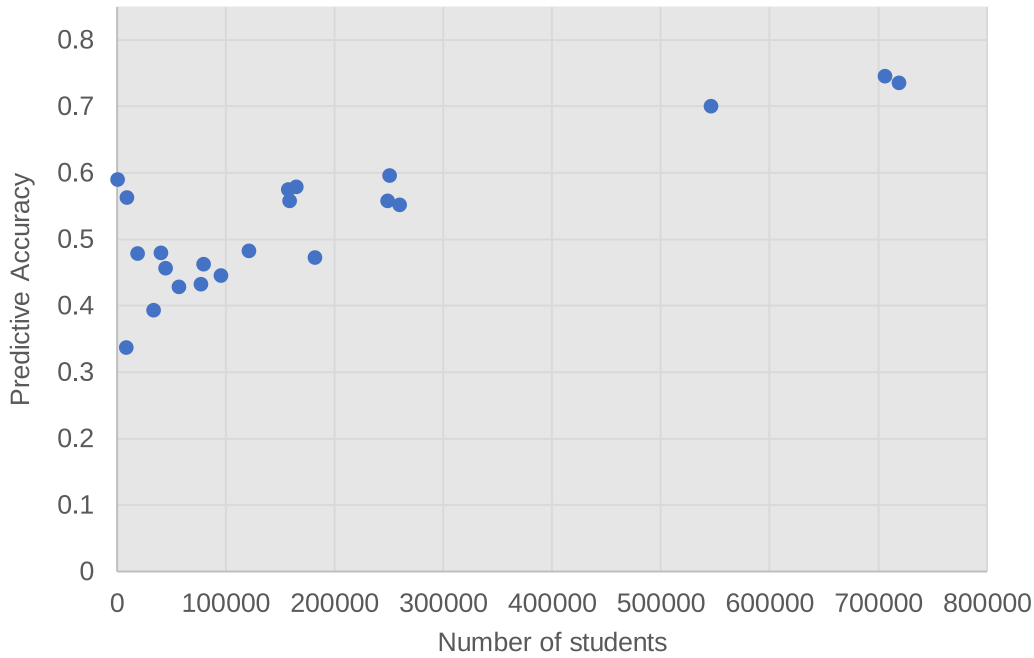


Figure 7.24 Plot of predictive accuracy by entry size for a range of GCSE subjects

To provide context for these measures, it is important that they are considered against analogous measures from a typical year. For the purposes of comparison, measures of qualification-level classification accuracy from a recent Ofqual study examining marking consistency[65] are presented in Figure 7.25. These metrics are based on the probability of examiners awarding marks to students' responses that result in the same grade being awarded as if the marking had been performed by the senior examiner. These probabilities of receiving the 'definitive' grade, therefore, provide a meaningful basis for comparison with the probabilities of an exact grade match included in the accuracy measures reported here.

To aid this comparison, Figure 7.25 includes the range of predictive accuracies reported above for the equivalent collections of subjects reported in the original plot. As can be seen, the range of predictive accuracy of the DCP approach is similar to the range of probabilities of the definitive grade being awarded during the standard marking and awarding process.

[65] Marking consistency metrics : an update

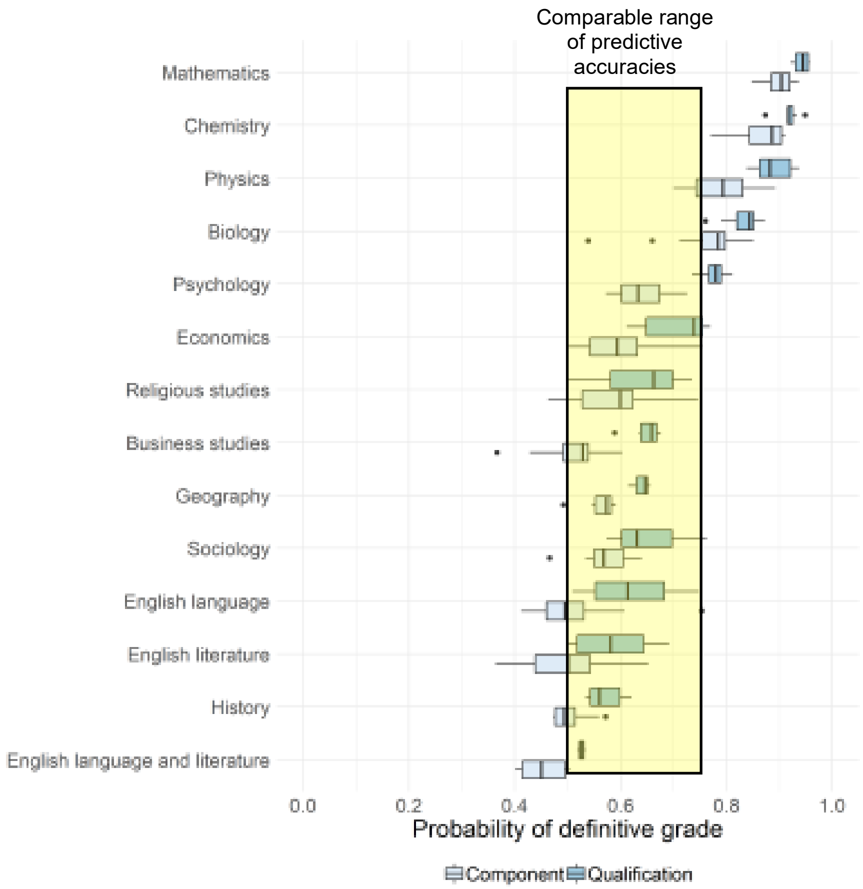


Figure 7.25 Probability of definitive grade being awarded based on an analysis of marking consistency reproduced from Ofqual 2018.

7.7 Summary

This section has outlined the testing conducted (including equalities analyses) to identify the most appropriate approach to realise standardisation this summer. The DCP approach was selected based on predictive performance, equity of treatment of centres with differing proportions of students from different groups and the practicalities of delivery.

Analyses have also been presented showing the predictive accuracy of the selected approach. This has shown levels of accuracy that are broadly comparable with the analogous measures from a typical year, when assuming that the rank orders submitted by centres are correct.

Beyond the absolute values of accuracy, it is important to also reflect on the nature of any inaccuracy. Any uncertainty in the outcomes produced through this method are unbiased and, the disassociation from CAGs, removes any risk of benefiting or disadvantaging centres that have taken different approaches. This contrasts with alternative approaches that may put greater emphasis on the CAGs. Not only are those approaches more susceptible to potential bias, it is not possible to validate the accuracy of these approaches due to the absence of equivalent authentic data for the purposes of testing.

While it is always desirable to maximise the predictive accuracy where possible, this approach ensures fairness through equity of treatment. The approach has also been proven not to differentially advantage or disadvantage any particular demographic or socio-economic group.

8 The direct centre-level performance approach

As identified in the Section 7, the DCP approach has been shown to provide the most accurate predictions for the purposes of standardising grades this summer. Section 6.2.3 provided a high-level description of the approach.

The purpose of this section of the report is to describe, in detail, the steps performed to implement the approach. This includes the arrangements made between exam boards to effectively operationalise the method, the core standardisation process itself and the handling of any qualification-specific issues (such as tiering), centre-specific issues (such as those with a small entry in a subject) and student-specific issues (such as the arrangements for private candidates).

The process described here is brought into effect by exam boards' compliance with Annex E of the regulatory requirements, introduced specifically for this summer[66].

8.1 Inter-exam board arrangements

A fundamental aspect of the arrangements this summer is the "national" approach necessary to be taken by exam boards. In a typical year, exam boards are distinguishable from one another by the assessments they deliver, but follow a common approach to the setting and maintenance of standards. A common approach to standard setting also exists this year (see Section 5), however, this also extends to the approach to standardisation. Not only is it important that all exam boards follow the same arrangements to ensure fairness to students across the cohort, but it is also important to ensure that the information used in the process is likewise common and defined at the national level.

To simplify the description that follows, issues relating to the interaction required between exam boards as part of the process are discussed separately here.

As described in Section 7, the approaches to standardisation rely on analyses being performed for each centre in each subject. A challenge to delivering this effectively is how centres distribute their entries across exam boards. Teachers have a range of motivations for deciding to enter their students for a particular specification offered in a subject. These include the specifics of the subject content covered by the specification and the appropriateness for their cohort of students, the approach to assessment and/or the supporting services offered by the exam board[67]. While drivers for this choice may remain constant, the profile of students within a centre or the teacher proficiencies or experiences of working with a specification can change over time. This may lead to centres switching their entry from one exam board to another between years. Also, in a significant minority of cases, centres choose to split their entry in a subject between specifications offered by different exam boards.

Shown in Table 8.1 is a breakdown of centre-subject combinations by qualification type this year – this is the number of instances of a centre entering at least one student in one subject. For example, 2 centres both deciding to enter students for

66 Requirements for the calculation of results in summer 2020
67 School and College Purchasing Behaviours

A level biology, chemistry and physics would contribute 6 centre-subject combinations. The table shows the number of centre-subject combinations where the centre has decided to split their entry across exam boards for a single subject and the number of students entering for a subject with a centre where the entry has been split. Where a centre splits their entry across specifications, this is recorded as a single centre-subject combination.

Table 8.1 Number and percentage of the entry at each qualification level which is split across exam-boards

	Total Centre-subject combinations	Centre-subject combinations with a split entry		Students entering for a subject-centre combination with a split entry	
		Number	Percentage	Number	Percentage
A level	52,736	1,732	3.3%	47,782	6.0%
AS	9,895	267	2.7%	3,601	4.5%
GCSE	87,607	2,558	2.9%	299,580	5.6%

While this is a relatively small proportion, the numbers of centres and students affected are significant.

A key consideration for the process this summer was what information centres taking this approach should submit to the process. More specifically, it was necessary to decide whether centres should submit a rank order at the subject level (therefore, providing a single rank order of students in the centre for a subject, irrespective of the exam board with which they entered) or whether they should submit a separate rank order for each specification for which they have entered. Taking the second of these approaches would have required statistical assumptions to be made about the distribution of students between specifications. Were it to be the case that students were randomly distributed between specifications/exam boards it may have been plausible to collect a separate rank order for each specification. This would have been undesirable from the perspective of statistical power; however, the risk of anomalous intra-centre adjustments would have been relatively low. The assumption that students are randomly distributed between exam boards would, however, be unreasonable given the evidence regarding the motivation for centres selecting different specifications. It is highly likely that centres have chosen to enter students for a particular specification for identifiable, non-random reasons. For example, in many instances where centres have split their entry, they are likely to have decided this split based on the relative ability of students. For this reason, centres were required to submit a single rank order for each subject for which they had entered students irrespective of how they were split between exam boards. We, therefore, required exam-boards to work together to co-ordinate the collection and processing of this information.

To ensure that students were treated consistently, irrespective of how their centre has chosen to enter students over time, a single national historical data set was constructed for use in the standardisation process. This ensures that students are neither advantaged nor disadvantaged due to historical entry decisions made by their centre.

Other points in the process where exam boards were required to put in place appropriate national arrangements through the standardisation process are specified in Annex E of the regulatory requirements for this summer[68]

The description of the approach to standardisation that follows, for simplicity, is articulated in the context of a centre that enters all the students for a subject with a single exam board. It also refers to the standardisation of a single subject, unless otherwise stated.

8.2 The core DCP approach

8.2.1 Step 1: Determine the historical grade distribution for the centre in the subject

The first step in applying the DCP approach is to identify the historical performance of students for each centre entering students in the subject this summer. This provides a start point for the calculation of each centre's predicted distribution of grades to be used this summer.

The historical performance of the centre is defined as the grade distribution achieved by students entering through that centre over recent years. This is articulated in terms of cumulative percentage of students achieving each grade or, more simply, the percentage of students achieving each grade or higher[69].

Throughout this description, reference is made to cumulative percentage grade distributions. These may be distributions of final calculated grades, grades at an intermediate step of the process or CAGs. For a grade-set 0 to M where 0 represents ungraded and M represents the highest grade for a qualification, the cumulative proportion grade distribution D, is defined as:

$$D = \{D_M \dots D_k \dots D_0\}$$

where:

- $D_k = \frac{n_k}{N}$,
- n_k is the number of students in the specified population allocated grade k or higher,
- N is the number of students in the specified population

When performing this step of the process, the number of years over which the historical data are aggregated is dependent on the qualification. This was considered in Section 7.2. For the purposes of later reference, the historical grade distribution created here is notated as $c_j = \{c_{Mj} \dots c_{kj} \dots c_{k0}\}$, where c_{kj} is the percentage of students at centre j achieving a grade k or higher.

68 Requirements for the calculation of results in summer 2020

69 Note that reference throughout this section is made to percentages or cumulative percentages, however, for the purposes of implementation and for simplification of the notation presented here, this is expressed mathematically as proportions or cumulative proportions.

As an example, a centre with the following set of grades, entering for a qualification for which the historical data are defined over three years, would have a historical grade distribution as defined below:

Year	A*	A	B	C	D	E	U	Total students
	Number of students achieving grade							
2017	2	5	7	9	4	2	1	30
2018	3	3	7	11	5	2	1	32
2019	1	3	7	8	4	2	2	27
Total	6	11	21	28	13	6	4	89
Total getting grade or better	6	17	38	66	79	85	89	89

$$c_j = \begin{Bmatrix} c_{A*j} = 6.7\%, c_{Aj} = 19.1\%, c_{Bj} = 42.7\%, c_{Cj} = 74.2\%, c_{Dj} = 88.8\%, \\ c_{Ej} = 95.5\%, c_{Uj} = 100.0\% \end{Bmatrix}$$

8.2.2 Step 2: Determine the historical value-added relationship for the subject

Having established the historical grade distribution, attention now turns to adjusting that distribution, if necessary, based on the prior attainment – a proxy for the general ability – of the cohort entering for the subject in each centre.

The key piece of information used for this purpose is the prior-attainment of students. The prior-attainment of groups of students is commonly used as either a control variable in experimental or operational research to account for potential differences in the relative ability of groups of students (e.g. see Ofqual, 2019[70]). It is also used operationally by exam boards to set and maintain qualification standards[71,72,73,74],

70 Ofqual (2019). *A level maths: Maintenance of Standards Investigation: Technical Report*. Coventry: Ofqual 19/6567/2. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851737/A-level_Maths_Technical_Report_Dec_2019_-_FINAL1965671.pdf

71 Taylor, R. & Opposs, D. (2018). Standard Setting in England: A levels. In *Examination Standards: How measures and meanings differ around the world*, ed. J-A Baird, T. Isaacs, D. Opposs, and L. Gray. London: UCL IOE Press.

72 Benton, T. & Lin, Y. (2011). *Investigating the relationship between A level results and prior attainment at GCSE*. Coventry: Ofqual. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/605906/2011-09-29-investigating-the-relationship-between-a-level-results-and-prior-attainment-at-gcse.pdf

73 Robinson, C. (2007). Awarding examination grades: current processes and their evolution. In *Techniques for monitoring the comparability of examination standards*, ed. P.E. Newton, J.-A. Baird, H. Goldstein, H. Patrick, and P. Tymms, 97–123. London: Qualifications and Curriculum Authority.

74 Baird, J-A. & Eason, S. (2004). *Statistical screening procedures to investigate inter-awarding body comparability*. London: Joint Council for Qualifications.

allowing differences in the ability of a cohort to be factored in. This was described in Section 5.2 in the context of cohort-level statistical predictions.

The prior attainment measure used for students sitting GCSE qualifications is their KS2 results. For AS, A level, EPQ and AEA students, the prior attainment is defined as their mean GCSE score. The definition of a prior-attainment matched student, for the purposes of standardisation, followed established practice. This means the student must be of target age for the qualification (16-years-old for GCSE, 17-years-old for AS, 18-years-old for A level, EQP and AEA) on the 31 August[75] in the relevant academic year, have a valid prior attainment record and, where mean GCSE score is used as the predictor, have results in a least 3 specifications at the age of 16.

It is important to note that neither through the standardisation process applied this year nor the use of cohort level predictions in a typical year, does a student's individual prior attainment dictate their outcome in a subject. Measures of prior-attainment are only used to characterise and, therefore, predict for group relationships between students.

Once the measure of prior-attainment has been established, to use it for predictive purposes, it is necessary to form a model of the expected value-added relationship; that being the relationship between the prior-attainment distribution of a cohort and the distribution of results achieved by a previous cohort in a particular qualification. Prediction matrices are a well-established method of articulating this type of relationship[71,72,73,74] and are used within the standardisation model to articulate value-added.

In simple terms, a prediction matrix is a cross-tabulation of student prior-attainment and the grade those students achieved in the qualification of interest, expressed in terms of cumulative percentages.

The first stage of the process to create a prediction matrix is to identify the cut-points that will be used to divide students into prior-attainment categories. This is achieved by performing the following two steps:

i. Identify all students in the historical data-set across all subjects at that qualification level who:
   - have the relevant prior-attainment measure,
   - meet any other selection criteria (such as age), and
   - achieved a valid result (a grade for the qualification or ungraded) in any subject.

   Students across AS, A level, EQP and AEA qualifications are identified for the purposes of setting cut-points for these qualifications. For GCSE, only students' historical results in GCSE qualifications are identified.
ii. Based on the prior-attainment distribution of students identified in step i, for each qualification level, determine the cut-points that divide the cohort into deciles. These cut-points define the points on the scale where students are split into different prior-attainment categories.

[75] 30 June for students in Northern Ireland

Once the prior attainment cut-offs have been identified the prediction matrix can be built through the following steps:

i. Identify prior-attainment matched students in the historical data that have a result in the subject of interest. Identify the prior attainment category into which each of these students fall such that:

$$A_i = \begin{cases} 1, & \text{if } \alpha_i \geq x_1 \\ d, & \text{if } x_d \leq \alpha_i < x_{d-1} \\ 10, & \text{if } \alpha_i < x_9 \end{cases}$$

where A_i is the prior attainment category for student i, with prior attainment value, α_i, compared to the prior-attainment decile cut-offs, $\{x\}$ where category 1 is the highest.

ii. Cross-tabulate these prior-attainment categories with the grades achieved by the students in the subject of interest. An example is shown below:

Prior-attainment category	Number of students in category achieving each grade						
	A*	A	B	C	D	E	U
1	5571	6564	2467	685	127	26	4
2	1352	5263	4864	2368	780	128	19
3	413	3002	4912	4122	1860	480	51
4	135	1394	3487	3991	2551	747	135
5	55	684	2169	3603	3086	1234	231
6	23	341	1245	2791	3142	1661	406
7	6	63	360	1024	1374	935	262
8	6	78	433	1242	2007	1648	601
9	1	26	128	453	890	982	538
10	1	12	36	123	272	357	404

iii. For each prior-attainment category, convert the grade distribution within that category to a cumulative percentage distribution. This represents the probability of a student with that level of prior-attainment achieving each grade or higher in the subject. This is performed below for the values given in step ii.

Prior-attainment category	Percentage of students in each category achieving each grade or better						
	A*	A	B	C	D	E	U
1	36.1	78.6	94.5	99.0	99.8	100.0	100.0
2	9.2	44.8	77.7	93.7	99.0	99.9	100.0
3	2.8	23.0	56.1	83.9	96.4	99.7	100.0
4	1.1	12.3	40.3	72.4	92.9	98.9	100.0
5	0.5	6.7	26.3	58.9	86.8	97.9	100.0
6	0.2	3.8	16.7	45.8	78.5	95.8	100.0
7	0.1	1.7	10.7	36.1	70.3	93.5	100.0
8	0.1	1.4	8.6	29.2	62.6	90.0	100.0
9	0.0	0.9	5.1	20.1	49.6	82.2	100.0

10	0.1	1.1	4.1	14.3	36.8	66.5	100.0

The result of step iii. is the prediction matrix reflecting the historical value-added relationship.

8.2.3 Step 3: Generate a historical 'prediction'

Once the cohort-level value added relationship has been established, it can be used to calculate the appropriate adjustment required for each centre entering for the subject. This adjustment is made to reflect the potential differences in ability of the cohort entering for the subject with a centre this year, compared to those entering with the same centre in previous years.

The first step in this process, to provide a basis for comparison, is to generate a historical prediction for each centre entering the subject based on the prior-attainment distribution of their historical cohort. This procedure is similar to the process that is performed at overall cohort-level due to the cohort entering from that centre being stronger or weaker based on their prior attainment profile.

To do this, the following process is followed:

i. Identify all students in the historical data set with the relevant prior attainment measure and meeting any other selection criteria entered for the subject with each centre.

ii. For each centre, split the prior-attainment matched students into categories based on the prior-attainment cut-points identified in Section 8.2.2, such that:

$$A_i = \begin{cases} 1, & \text{if } \alpha_i \geq x_1 \\ d, & \text{if } x_d \leq \alpha_i < x_{d-1} \\ 10, & \text{if } \alpha_i < x_9 \end{cases}$$

using the same notation as defined above.

iii. For each prior attainment category for each centre, determine the (non-integer) number of students who would have been predicted to achieve each grade. This is achieved by multiplying the number of students in the prior-attainment category by the probability of students in that category achieving each grade or better (as defined by the prediction-matrix calculated above).

iv. For each grade, sum the number of students across the prior-attainment categories predicted to achieve each grade or higher.

Steps iii. and iv. are performed below using the example of the centre introduced in Section 8.2.1. It should be noted that the number of students quoted below differs from that presented in Section 8.2.1 due to some students not having measures of prior-attainment.

Students in prior-attainment category	Number of historical students predicted to achieve each grade or better in each prior-attainment category for the centre						
	A*	A	B	C	D	E	U
9	3.2	7.1	8.5	8.9	9.0	9.0	9.0
7	0.6	3.1	5.4	6.6	6.9	7.0	7.0

10	0.3	2.3	5.6	8.4	9.6	10.0	10.0
16	0.2	2.0	6.5	11.6	14.9	15.8	16.0
11	0.1	0.7	2.9	6.5	9.5	10.8	11.0
12	0.0	0.5	2.0	5.5	9.4	11.5	12.0
10	0.0	0.2	1.1	3.6	7.0	9.3	10.0
3	0.0	0.0	0.3	0.9	1.9	2.7	3.0
5	0.0	0.0	0.3	1.0	2.5	4.1	5.0
2	0.0	0.0	0.1	0.3	0.7	1.3	2.0
Cumulative students	4.4	15.9	32.6	53.2	71.5	81.5	85.0

v. Expressing the sum of students predicted to achieve each grade or higher provides the historical predicted grade distribution for that centre, had they followed the national value-added relationship. This distribution is defined as $p_j = \{p_{Mj} \dots p_{kj} \dots p_{0j}\}$ for later reference. Using the example above, this results in:

$$p_j = \begin{Bmatrix} p_{A*j} = 5.2\%, p_{Aj} = 18.8\%, p_{Bj} = 38.3\%, p_{Cj} = 62.6\%, p_{Dj} = 84.1\%, \\ p_{Ej} = 95.9\%, p_{Uj} = 100.0\% \end{Bmatrix}$$

It should be noted that this step in the process generates a prediction assuming that the centre performed in line with the national value-added relationship for that subject. This is unlikely to be the case for most centres in most subjects. Indeed, it can be seen from the comparison of p_j and c_j (ignoring the presence of non-prior-attainment matched students in c_j) that the centre in the example provided here out-performed the national value-added relationship. Due to the differential nature of the calculations that follow, these effects are removed when forming the actual prediction for the centre this year.

8.2.4 Step 4: Generate the initial prediction for the current students

The next step follows the same procedure as that described in Step 3, however, this time, it is performed for the current cohort entering for the subject with each centre. The result of this process is the predicted grade distribution for the centre in the subject this year, assuming the centre would have followed the national value-added relationship.

To generate this prediction, the steps ii. to v. above are repeated for prior-attainment matched students entering for the subject with the centre in summer 2020. This is exemplified below, using the same centre used above.

Students in prior-attainment category	Number of historical students predicted to achieve each grade or better in each prior-attainment category for the centre						
	A*	A	B	C	D	E	U
4	1.4	3.1	3.8	4.0	4.0	4.0	4.0
4	0.4	1.8	3.1	3.7	4.0	4.0	4.0
3	0.1	0.7	1.7	2.5	2.9	3.0	3.0
6	0.1	0.7	2.4	4.3	5.6	5.9	6.0
5	0.0	0.3	1.3	2.9	4.3	4.9	5.0

6	0.0	0.2	1.0	2.7	4.7	5.7	6.0
1	0.0	0.0	0.1	0.4	0.7	0.9	1.0
1	0.0	0.0	0.1	0.3	0.6	0.9	1.0
1	0.0	0.0	0.1	0.2	0.5	0.8	1.0
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Cumulative students		2.0	7.0	13.6	21.1	27.3	30.2	31.0

The resulting predicted grade distribution for this year's cohort, had they followed the national value-added relationship, is denoted as $q_j = \{q_{Mj} \ldots q_{kj} \ldots q_{0j}\}$. In this example, it follows that:

$$q_j = \begin{Bmatrix} q_{A*j} = 6.5\%, q_{Aj} = 22.5\%, q_{Bj} = 43.7\%, q_{Cj} = 68.1\%, q_{Dj} = 88.0\%, \\ q_{Ej} = 97.5\%, q_{Uj} = 100.0\% \end{Bmatrix}$$

8.2.5 Step 5: Determine the lowest prior-attainment match-rate for each centre

As in the example considered above, not all students entering for the subject will have prior-attainment data. This may be because not all students sat the relevant assessments or, for practical data-matching reasons, it has not been possible to reliably link the student back to their prior-attainment measure. For information, presented in Annex D is a breakdown of the prior-attainment match rates for each qualification based on the summer 2020 entry data.

Steps 2 to 4, described in Sections 8.2.2 to 8.2.4 are included in the process in order to reflect potential changes in the prior-attainment profile of students entering for a subject with each centre. However, if a centre were to have a low-proportion of prior-attainment matched students, it would be inappropriate for differences in the prior attainment profile for this sub-set of students to strongly influence this year's prediction. In contrast, where the match rate is high in both the current year and in the historical data, it is more appropriate for differences in the prior-attainment profile over time to influence the predicted outcome for the centre.

To determine the influence that the prior-attainment information should have over the centre-level prediction in the subject for this summer it is necessary to consider the match rate this year and in the historical data. A summary measure of this match rate across years is defined as:

$$r_j = \min\left(\frac{n'_{\text{hist},j}}{n_{\text{hist},j}}, \frac{n'_{\text{cur},j}}{n_{\text{cur},j}}\right)$$

where $n_{\text{hist},j}$ is the total number of students in the historical data for the centre entering the subject, $n'_{\text{hist},j}$ is the equivalent number for prior-attainment matched students and $n_{\text{cur},j}$ and $n'_{\text{cur},j}$ are corresponding measures for the current year, respectively. For the example presented above, the following figures result:

Year	All students	Prior-attainment matched students		
2017	30	28	$n_{\text{hist},j}$ = 91	$n'_{\text{hist},j}$ = 87
2018	32	32		
2019	29	27		
2020	35	31	$n_{\text{cur},j}$= 35	$n'_{\text{cur},j}$ = 31

$$r_j = \min\left(\frac{87}{91}, \frac{31}{35}\right)$$

$$= \min(0.96, 0.89)$$

$$= 0.89$$

8.2.6 Step 6: Determine the centre-level predicted grade distribution for summer 2020

To form this year's predicted-grade distribution for each centre in the subject, the historical information calculated in Step 1 is adjusted based on the predictions resulting from Steps 3 and 4 and the weight that should be given to this adjustment as determined by the match-rate calculated in Step 5.

This adjusted prediction $P_j = \{P_{Mj} \dots P_{kj} \dots P_{0j}\}$ is calculated as:

$$P_{kj} = (1 - r_j)c_k + r_j(c_{kj} + q_{kj} - p_{kj})$$

$$= c_{kj} + r_j(q_{kj} - p_{kj})$$

The equation can be easily broken down to explain its operation:

$$P_{kj} = (1 - r_j)c_k + r_j(c_{kj} + q_{kj} - p_{kj})$$

- c_k: Unadjusted historical outcome
- r_j (second term): Proportion of matched students. This scales the influence the adjusted historical outcome has over the final prediction.
- $q_{kj} - p_{kj}$: Difference in outcome due to change in prior attainment
- $(1 - r_j)$: Proportion of unmatched students. This scales the influence of the unadjusted historical outcome has over final prediction.
- $(c_{kj} + q_{kj} - p_{kj})$: Historical outcome with the full adjustment due to the difference in prior-attainment

As can be seen above, the first term in the equation above controls the amount of influence the raw historical outcome has over the prediction at the grade, based on the proportion of unmatched students; if the lowest match rate (r_j) across the historical and current years was 60%, this term would contribute 40% of the weight to the prediction. The second term in the equation controls the influence of the prior-attainment adjusted outcome; so, in this brief example would contribute 60% of the weight.

From this is can be easily seen that, in a situation where a centre has no prior attainment matched students, the centre-level prediction is defined entirely by the historical centre outcome since $r_j = 0$ leading to the second term collapsing to zero resulting in $\{P_j\} = \{c_j\}$.

The output from this step is the centre-level prediction for the subject.

Using the example considered above, this shows that the increase in the prior attainment profile for students in summer 2020 relative to previous years gives rise to an increase in the centre-level prediction for the subject as shown below[76]:

76 Note that for illustrative clarity only, the calculations shown here are based on rounded figures.

$$P_j = \begin{bmatrix} P_{A*j} \\ P_{Aj} \\ P_{Bj} \\ P_{Cj} \\ P_{Dj} \\ P_{Ej} \\ P_{Uj} \end{bmatrix} = \begin{bmatrix} c_{A*j} \\ c_{Aj} \\ c_{Bj} \\ c_{Cj} \\ c_{Dj} \\ c_{Ej} \\ c_{Uj} \end{bmatrix} + r_j \left(\begin{bmatrix} q_{A*j} \\ q_{Aj} \\ q_{Bj} \\ q_{Cj} \\ q_{Dj} \\ q_{Ej} \\ q_{Uj} \end{bmatrix} - \begin{bmatrix} p_{A*j} \\ p_{Aj} \\ p_{Bj} \\ p_{Cj} \\ p_{Dj} \\ p_{Ej} \\ p_{Uj} \end{bmatrix} \right)$$

$$= \begin{bmatrix} 6.7 \\ 19.7 \\ 42.7 \\ 74.2 \\ 88.8 \\ 95.5 \\ 100.0 \end{bmatrix} + 0.89 \left(\begin{bmatrix} 6.5 \\ 22.2 \\ 43.7 \\ 68.1 \\ 88.0 \\ 97.5 \\ 100.0 \end{bmatrix} - \begin{bmatrix} 5.2 \\ 18.8 \\ 38.3 \\ 62.6 \\ 84.1 \\ 95.9 \\ 100.0 \end{bmatrix} \right) = \begin{bmatrix} 6.7 \\ 19.7 \\ 42.7 \\ 74.2 \\ 88.8 \\ 95.5 \\ 100.0 \end{bmatrix} + \begin{bmatrix} 1.2 \\ 3.0 \\ 4.8 \\ 4.9 \\ 3.5 \\ 1.4 \\ 0.0 \end{bmatrix}$$

$$= \begin{bmatrix} 7.9 \\ 22.7 \\ 47.5 \\ 79.1 \\ 92.3 \\ 96.9 \\ 100.0 \end{bmatrix}$$

8.2.7 Step 7: Determine the notional student grades for summer 2020

Having performed the steps outlined above, it is possible to produce notional grades for this summer's students. This is performed by overlaying the rank order provided by the centre onto each centre's predicted cumulative percentage grade distribution such that the proportion of students awarded each grade within the centre matches the predicted distribution as closely as possible.

It is noted that the rank orders of students collected from centres were articulated separately for each grade, as discussed in Section 2.3. For the purposes of this stage in the process, the rank order for each centre was restructured into a single contiguous rank order covering all grades. This restructuring retains the overall ordering of students within the centre.

The resulting distribution for the example given is provided below:

	A*	A	B	C	D	E	U
P_j	7.9%	22.7%	47.5%	79.1%	92.3%	96.9%	100.0%
Cumulative student outcome	3	8	17	28	32	34	35
Rank of students getting grade	1-3	4-8	9-17	18-28	29-32	33&34	35
Notional cumulative percentage distribution	8.6%	22.9%	48.6%	80.0%	91.4%	97.1%	100.0%

8.2.8 Step 8: Mark imputation

As discussed at length in Section 5, cohort-level statistical predictions play an important role in ensuring that the overall distribution of grades awarded this summer is appropriate and credible. Without use of these predictions, inadvertent severity or leniency could occur at subject level, to variable degrees across subjects, potentially advantaging or disadvantaging students across the current year or across previous/latter years.

To enable these adjustments to overall outcomes to be made, it is necessary to transform the information generated through the process described above into imputed marks. The aim of this step of the process is to identify the students who would be most likely to move up or down a grade, where the adjustments to the overall outcomes are necessary.

To enable this process, a mark scale with notional cut-scores is constructed. The length of the mark scale is arbitrary as the marks imputed through the calculation are continuous (fractional) numbers rather than integers relating to a discrete mark scale, as would conventionally be the case. For the purposes of simplicity, the mark scale is defined as having 100 marks per grade and ranging from 0 to $M \times 100$, where M is the number of grades available for the qualification (including ungraded). Notional cut-scores for each grade therefore occur at 100-mark intervals – for example, at A level: A* notional cut-score = 600, A = 500 etc.

To determine the imputed marks for students, the notional grade for each student (as determined above) is used in combination with their positions in both the rank order submitted by their centre and the predicted grade distribution, P_j.

To perform this imputation, the following calculation is performed:

$$X_{ij} = 100 \times \left(k_{ij} + \left(\frac{P_{k_{ij}j} - \frac{(\rho_{ij} - 0.5)}{n_{cur,j}}}{P_{k_{ij}j} - P_{(k_{ij}+1)j}} \right) \right)$$

Where X_{ij} is the imputed mark for the ith student in the jth centre, ρ_{ij} is the overall rank position of the student in the centre for the subject, and k_{ij} is the largest possible grade satisfying the inequality:

$$\frac{(\rho_{ij} - 0.5)}{n_{cur,j}} \leq P_{kj}$$

The effect of applying this equation is, individually for each centre, to space students evenly across the mark range available for that notional grade. The mark for the lower most student at each notional grade is calculated dependent on the difference between the actual predicted grade distribution, P_j, and the notional grade distribution that it was possible to achieve due to the discrete nature of the students making up the actual distribution. Instances where the rounding is favourable and only just led to a student being awarded the higher grade means the lower most student has an imputed mark very close to the notional cut-score and vice versa. The consequences are that such that, when the notional cut-scores are adjusted to realise the cohort-level predictions, those students close to the notional cut-scores are prioritised in being regraded.

To aid further understanding, an annotated version of this expression is provided below:

$$X_{ij} = 100 \times \left(k_{ij} + \left(\frac{P_{k_{ij}j} - \frac{(\rho_{ij} - 0.5)}{n_{cur,j}}}{P_{k_{ij}j} - P_{(k_{ij}+1)j}} \right) \right)$$

Multiplier to create 100 marks per grade

Fraction reflects the position of the student within the notional grade based on their rank order and the prediction for the centre

Percentile rank of the student within the centre

Offset introduced to reflect notional grade awarded to the student

Percentage of students predicted to achieve the grade notionally awarded to the current student

To demonstrate the functioning of this expression, shown in Figure 8.1 is the value of the fractional term, plotted against the rank position of each student, calculated for the example used in this section. For illustrative purposes, the blue lines link students allocated the same grade at Step 7. This shows the spread of students across the mark range for that grade.

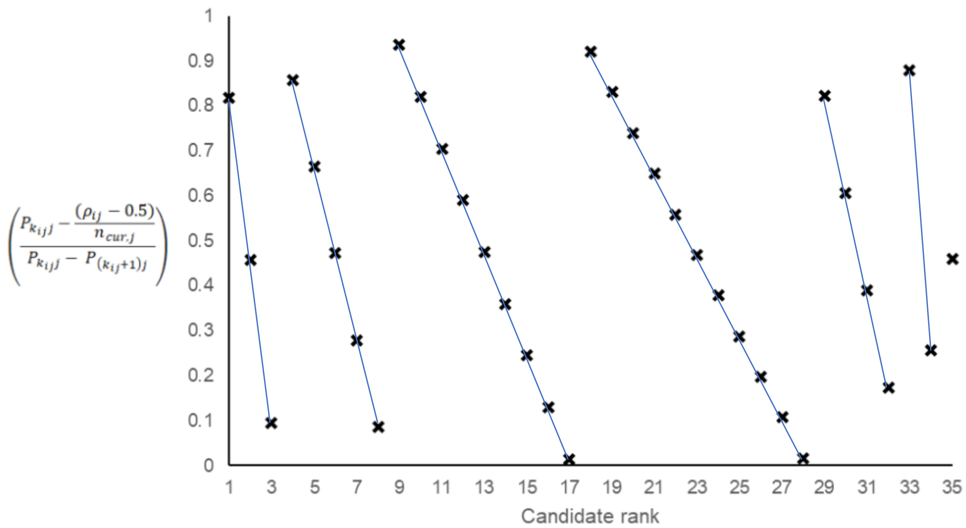


Figure 8.1 Fractional term of the imputed mark calculation equation based on example data

Combining the fractional term with the rest of the equation has the effect of adding a 100-mark offset for each grade, combining these slopes into a single distribution. This results in the imputed marks shown in Figure 8.2.a demonstrating the monotonic relationship that is formed between the rank order and marks. Figure

8.2.b demonstrates the positioning of the student closest to the grade B notional cut-score based on the proximity of the outcome compared to that predicted.

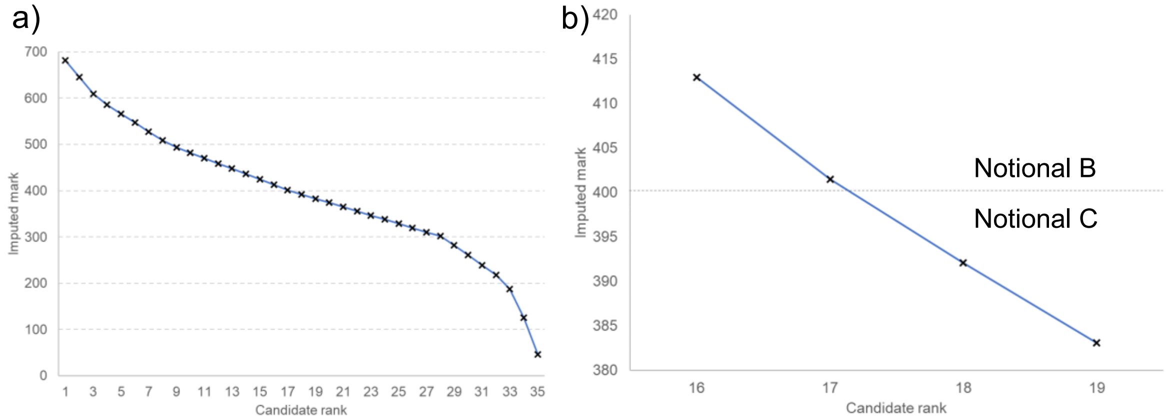


Figure 8.2 Imputed marks calculated for the example centre used for the purposes of description

8.2.9 Step 9: Cut-score setting

The final stage of the process is to set the cut-scores to achieve an appropriate overall standard. This determines students' calculated grades. This part of the process is similar to a key part of the grade boundary setting process[71,73] in a typical year. Usually, in preparation for an awarding meeting, exam boards identify the grade boundaries that would most closely represent a maintenance of statistical standards over time based on the cohort-level prediction. This enables senior examiners to be presented with an appropriate range of work to scrutinise when recommending the final grade boundaries. The process of identifying cut-scores based on imputed marks is analogous to this process of identifying statistically recommended boundaries with the key differences being: the mark data arise from the process described above rather than from the marking of student work, that the marks available are defined on a continuous rather than discrete scale and that no student work is subsequently available for scrutiny.

As in a typical year, this part of the process is performed using prior-attainment matched students only to ensure, as far as is possible, a like-for-like comparison of cohorts which takes into account any overall variation in prior attainment. The cut-scores are set at the imputed mark for the student in the cohort who most closely reflects the statistical prediction for the subject.

This year, to ensure a consistency of standard across exam boards, this process is performed nationally, meaning that a single prior-attainment matched student mark distribution is formed across exam boards with the cut-score being set against a single national prediction at each grade.

8.3 Private candidates

During the development of the standardisation approach, questions were raised regarding the handling of private candidates. Formally, a private candidate is any student for whom there is no ‘Relevant Centre’ which is defined in the GQ extraordinary regulatory framework[77] as:

> *Relevant Centre - In relation to a Learner, a Centre which –*
>
> *(a) has purchased the GQ Qualification on behalf of the Learner, and*
>
> *(b) materially contributed to the preparation of the Learner for the assessment (whether through teaching or instruction provided by Teachers employed by it or otherwise).*

In practice, there is a wide range of different local arrangements that are in place for private candidates, but, typically, a private candidate can be considered as one who does not have as close a relationship with the centre as other students in the centre. It may be that, in a typical year, the relationship between the student and the centre is purely functional with the student only engaging for the purposes of sitting the assessments themselves.

The questions raised regarding the handling of private candidates were both technical and behavioural.

From a technical perspective, the argument that private candidates should follow the historical behaviour of the centre (in terms of their absolute outcomes or their value-added) is questionable. This is due to their more distant relationship with the centre and the likelihood that the quality of teaching and learning at the centre will have had no or little impact on their performance, had the exams been sat this summer. In addition, it is likely that it would have been more challenging for the centres to form a good evaluation of the potential performance of a private candidate given the reduced familiarity. To strike a balance between enabling private candidates to continue to receive results this summer, in line with other students, yet ensure that the judgement of the ability of private candidates was sufficient, guidance was put in place to facilitate these judgements[78]. To provide assurances that the information submitted by centres regarding private candidates were legitimate this included clarification that “*Heads of centre must be as confident in the centre assessment grades and rank order for private candidates as they are for their other students*”.

In addition to these more technical considerations, questions were also raised regarding the potential behaviours which may take place within centres regarding the ranking of private candidates. The concern raised was that private candidates might be potentially disadvantaged through the process by them being inappropriately positioned lower in the rank order than their abilities would suggest. The claimed motivation for this would be to protect the interests of students with whom the centres had a more established relationship following any adjustment. This is aligned to some perceptions that, if a private candidate were placed at the top of, or high in, a centre’s rank order, they may achieve a grade in place of a non-private candidate.

To protect against these effects, during the process of centres submitting their CAGs and rank orders we announced that private candidates would not impact on the

[77] Extraordinary regulatory framework: General Qualifications, COVID-19 Conditions and Requirements

[78] 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

standardisation of other students within the centre[79]. As the CAGs of private candidates will have been subject to the same judgement process and therefore potential severity/leniency as other students entering the subject with the centre it is important that these CAGs are also standardised, but for the reasons given they are treated differently through the process.

This was operationalised by excluding private candidates from the process described in Section 8.2. This includes the removal of private candidates from the historical data for each centre and in the current cohort. However, to ensure that the CAGs of private candidates are standardised in line with the adjustment applied to the rest of students within the centre entering the subject, a process of post-hoc slotting has been applied. This process works by standardising students within the centre as described with private candidates omitted, then, at the end of the process, the grades for each private candidate are determined by slotting them back into the original rank order submitted by the centre. This post-hoc slotting was based on the following rules:

Scenario 1) If the next non-private candidate ranked[80] above the private candidate has the same grade, h, as the next non-private candidate ranked below the private candidate, the private candidate is allocated the same grade, h. This scenario is illustrated below.

Student	Calculated Grade	Rank
Student A	A	1
Student B	A	2
Student C	B	3
Student D	B	4
Student E	B	5
Student F	B	6
Student G	C	7
Student H	C	8
Student I	C	10
Student J	C	11
Student K	C	12
Student L	D	13
Student M	D	14

Private candidate

Students above and below have a calculated grade of C, therefore Student N's calculated grade is C

Student N Rank = 9

Scenario 2) If the next non-private candidate ranked above the private candidate has a grade, h, which is different to the next non-private candidate ranked below the private candidate with a grade l, the private candidate is allocated their CAG, grade provided it is in the range, h to l. If the private candidate's CAG is above, h, they should be

[79] Making grades as fair as they can be: advice for schools and colleges

[80] It is noted that the rank orders of students collected from centres were articulated separately for each grade. For the purposes of this stage in the process, and also illustratively here for the purposes of clarity, the rank order for each centre was restructured into a single contiguous rank order covering all grades. The restructuring retained the original ordering of students within the centre.

awarded h. If the private candidate's CAG is below, l, they should be awarded l.

Student	Calculated Grade	Rank
Student A	A	1
Student B	A	2
Student C	B	3
Student D	B	4
Student E	B	5
Student F	B	6
Student G	C	8
Student H	C	9
Student I	C	10
Student J	C	11
Student K	C	12
Student L	D	13
Student M	D	14

Private candidate

If Student N's CAG is a B or higher, the calculated grade is B. If Student N's CAG is C or lower, the calculated grade is C

Student N Rank = 7

Scenario 3) If the private candidate has no non-private candidates above them in the rank order, and the next non-private candidate ranked below them has a grade l, the private candidate will receive the CAG if it is higher than or equal to l. If the private candidate's CAG is lower than l, they will be awarded grade l.

Student	Calculated Grade	Rank
Student A	A	2
Student B	A	3
Student C	B	4
Student D	B	5
Student E	B	6
Student F	B	7
Student G	C	8
Student H	C	9
Student I	C	10
Student J	C	11
Student K	C	12
Student L	D	13
Student M	D	14

Private candidate

Student N Rank = 1

If Student N's CAG is an A, the calculated grade is an A*. If Student N's CAG is an A or lower, the calculated grade is A*

Scenario 4) If the private candidate has no non-private candidates below them in the rank order, and the next non-private candidate ranked above them has a grade h, the private candidate will receive the CAG if it is lower than or equal to h. If the private candidate's CAG is higher than h they will be awarded grade h.

Student	Calculated Grade	Rank
Student A	A	1
Student B	A	2
Student C	B	3
Student D	B	4
Student E	B	5
Student F	B	6
Student G	C	7
Student H	C	8
Student I	C	9
Student J	C	10
Student K	C	11
Student L	D	12
Student M	D	13

Private candidate

If Student N's CAG is a D or higher, the calculated grade is a D. If Student N's CAG is an E or lower, the calculated grade is the CAG

Student N Rank = 14

Exceptions to these arrangements are private candidates entering for a subject via an identified distance learning provider. By definition, all students entering through these centres are private candidates and, therefore, the presence of private candidates in this year's cohort would not be atypical. For these identified centres, the process of standardisation is conducted as described in Section 8.2 for all students.

8.4 Centres with a small entry in a subject

One of the principles on which the standardisation approach was based is that more weight should be placed on the statistical historical evidence of centre performance (given the prior attainment of students) than the submitted CAGs (see Section 4.2). The motivation for this decision is to ensure that any residual leniency or severity in the CAGs is removed equally across centres reducing inter-centre unfairness and to ensure that any overall leniency (which is likely to arise from a tendency towards leniency in the CAGs) is handled in a way that does not advantage or disadvantage centres who have taken different approaches. The consequence of this decision is that, in the majority of cases, the statistical model will be applied to determine the grade distribution for each centre in each subject.

An intended exception to this situation was for centres with a small number of entries in a subject. This reflects the weakening of the statistical evidence as the number of students reduces. In these circumstances, it is necessary to move from the statistical evidence as the primary source of evidence to the CAGs.

Discussions regarding the size of a centre's entry in a subject and the relevance of any statistical approach in those circumstances are frequently positioned in terms of centres with a small entry in the subject in the current year. Consideration must also be given, however, to those centres that have had a small entry in the subject across the years on which the centre-level prediction is based, as this also weakens the statistical evidence. Centres with a small entry in 2020 and those with a small entry across the historical data are, therefore, considered together for the purposes of discussion here and for the purposes of the standardisation model.

One approach to determining the balance between the statistical evidence and the CAGs would be to draw directly on the standard errors in the statistical estimation process and weight the influence of the CAGs/statistical outcomes on that basis. A potential form of the relationship is shown graphically in Figure 8.3 and algebraically below[81].

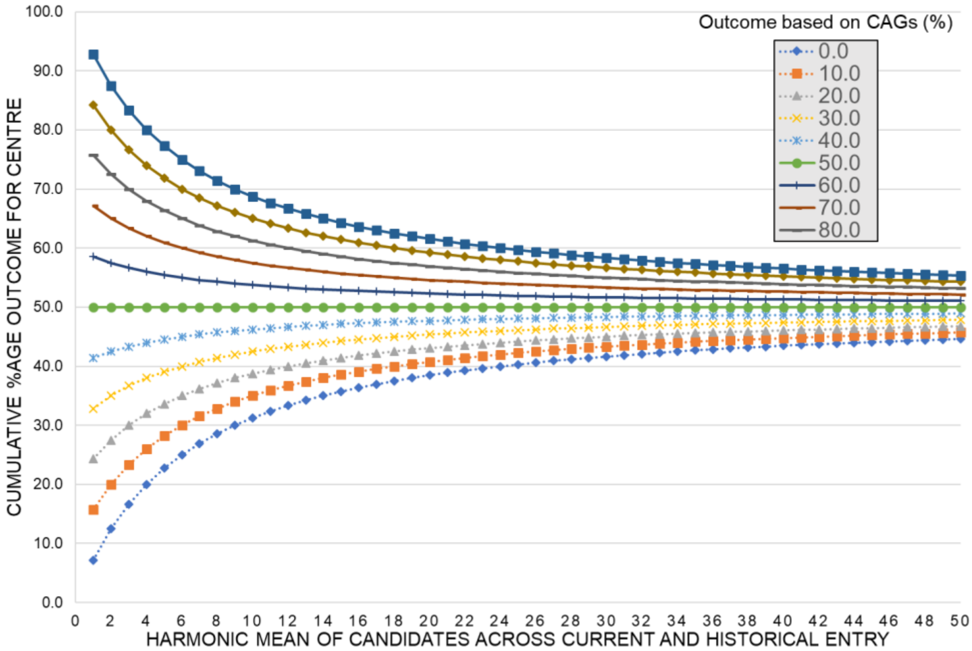


Figure 8.3 Potential weighting of CAGs based on current and historical entry size linked to standard error

$$P'_{kj} = (\bar{n}_{Hj} P_{kj} + 6 f_{kj}) / (\bar{n}_{Hj} + 6)$$

$$\bar{n}_{Hj} = \left(\frac{0.5}{n_{\mathrm{cur},j}} + \frac{0.5}{n_{\mathrm{hist},j}} \right)^{-1}$$

where $\bar{n}_{Hj}$ is the harmonic mean of the centre entry across the current year ($n_{\mathrm{cur},j}$) and the historical years ($n_{\mathrm{hist},j}$), P'_{kj} the revised prediction of the centre-level grade distribution, P_{kj} is the centre-level predicted grade distribution as defined by the statistical model (see Section 8.2.6) and f_{kj} is centre-level grade distribution as defined by the CAGs. See Annex H for an explanation and exemplification of the functioning of harmonic mean.

The example provided in Figure 8.3 is for a statistically predicted outcome of 50% at the chosen grade for a range of outcomes defined by the CAGs. For example, the dark blue line with square markers is for the case where the CAGs correspond to an outcome of 80% compared to the 50% indicated by the statistics. This relationship shows the increased weight this approach places on the CAGs as the number of students reduces. It is also important to note the role of the harmonic mean which considers the number of students in both the current year and in the historical data.

81 Benton, T. Personal communication, June 2020.

A significant challenge with taking this approach is that this relationship is asymptotic to the outcome that would be delivered by the statistical model; even for relatively large centre sizes, a notable difference exists between the statistically predicted outcomes and those weighted by the CAGs.

The consequence of this effect is that, to retain the principle of prioritising the statistical evidence for 'larger' centres, it is necessary to identify a discrete point on the entry size axis at which the transition from the CAG weighted prediction to the pure statistical prediction. This would result in a relationship of the form shown in Figure 8.4.

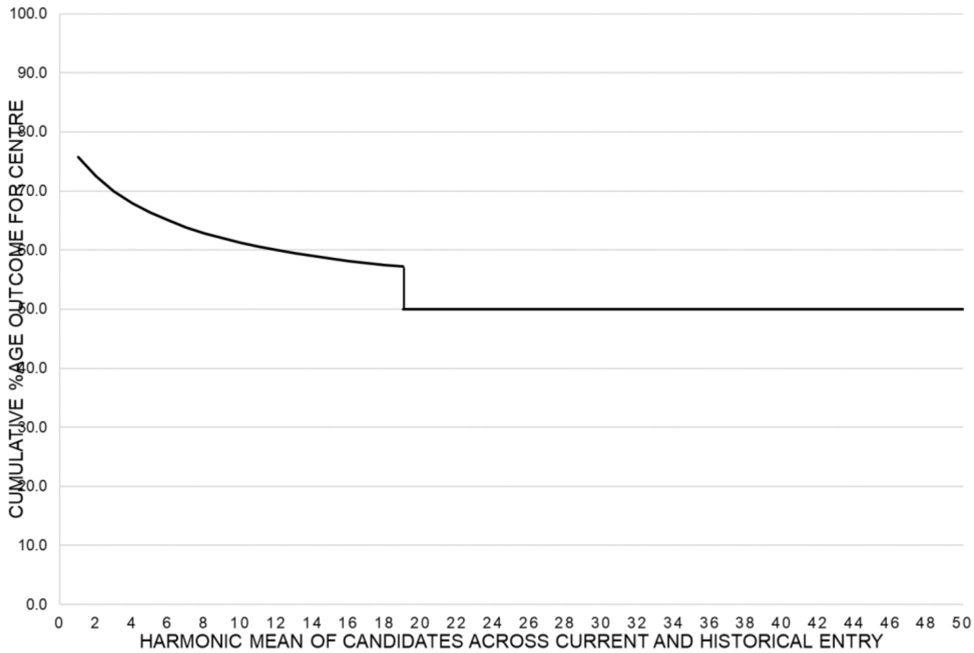


Figure 8.4. Sharp transition between the continuous relationship described above for small centres and use of the statistical evidence only

Having a sharp transition between the use of approaches is problematic given the differential treatment for centres either side of this transition. To approximate this more theoretically grounded relationship in a way that avoids a sharp transition in the relationship and provides additional protection against over-interpretation of the statistical data for students from centres with the smallest entries, the approach illustrated in Figure 8.5, and specified algebraically below, is applied.

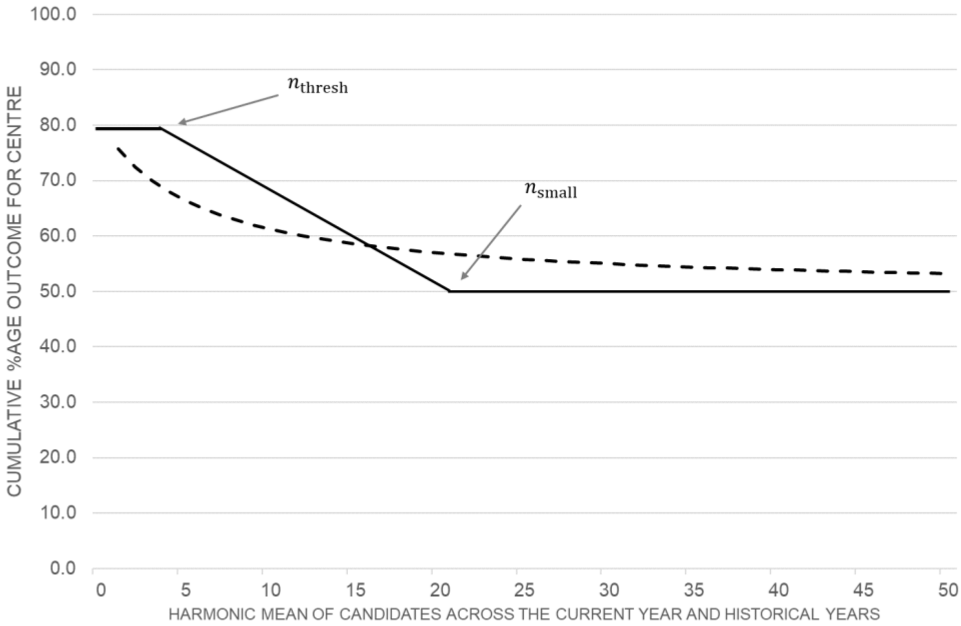


Figure 8.5. Proposed form of the relationship weighting CAGs against the statistical evidence for small centres

With the revised centre-level grade distribution, P'_{kj}, defined as:

$$P'_{kj} = \begin{cases} f_{kj}, \text{ if } \bar{n}_{Hj} < n_{\text{thresh}} \\ f_{kj} \dfrac{(n_{\text{small}} - \bar{n}_{Hj})}{(n_{\text{small}} - n_{\text{thresh}})} + P_{kj}\left(1 - \dfrac{(n_{\text{small}} - \bar{n}_{Hj})}{(n_{\text{small}} - n_{\text{thresh}})}\right), \text{ if } n_{\text{thresh}} \leq \bar{n}_H \leq n_{\text{small}} \\ P_{kj}, \text{ if } \bar{n}_{Hj} > n_{\text{small}} \end{cases}$$

While this relationship leads to the introduction of two thresholds, the transition from one region of the relationship to another (later referred to as the ‘taper’) is without a sharp discontinuity in treatment of centres with marginally different entry sizes.

The motivation for the constant region where the centre’s grade distribution follows the CAGs is to avoid anomalous and potentially indefensible adjustments for centres with a particularly small entry.

8.4.1 Standardisation of centres without historical data

Due to the necessary reliance on historical data to predict the outcomes for centres this summer, trying to predict outcomes for centres without any historical data is problematic. Approaches considered to address this issue were for the historical data for such a centre to be replaced with either the national grade distribution for that subject or for a similar, segmented, approach to be explored where the centre’s history is modelled based on properties such as centre type and 2020 prior-attainment profile.

Given the discussion above regarding centres with a small entry (including those with a small historical entry) it is clear, however, that centres with no historical data at all are a limiting case of this scenario. To take an approach to defining the historical data for a centre such as that described above would potentially lead to a notably different treatment of a centre with no historical data compared to another centre from which a single student had entered during the period from which historical data are being drawn. This would be potentially unfair and indefensible. These centres are, therefore, considered as having $\bar{n}_H$ = 0 leading to $P'_{kj} = f_{kj}$ with students at these centres being awarded their CAGs.

8.4.2 Handling of potential overall leniency

Based on the research literature, when designing the standardisation approach, it was anticipated that, overall, the CAGs submitted by centres would have a tendency towards leniency.[82,83,84,85,86,87,88] As the proposed approach outlined above involves the CAGs playing a greater role in the calculated grades for centres with a small entry in a subject, it was, therefore, anticipated that is likely that this would introduce an upward pressure on overall outcomes. There were two approaches that were considered to handle this likely leniency in overall outcomes:

> **Option 1)** Adjust the overall outcomes to remove the leniency effects of putting greater weight on the CAGs for centres with a small entry. This would mean overall outcomes in all subjects would closely meet the cohort-level statistical prediction.
>
> **Option 2)** Allow deviation from the statistical predictions to occur due the greater weighting on the CAGs for centres with a small entry.

82 Dhillon, D. (2005). Teachers' estimates of students' grades: Curriculum 2000 Advances Level Qualifications. *British Educational Research Journal, 31*(1), 69-88.

83 Department for Education & Skills (2005). *Improving the higher education applications process: A consultation paper.* Retrieved from: https://dera.ioe.ac.uk/5564/1/Improving%20the%20HE%20Applications%20Process%20-%20consultation%20paper%20%28PDF%29.pdf

84 Gill, T. & Rushton, N. (2011). *The accuracy of forecast grades for OCR A levels. Statistics Report Series No.26.* Cambridge Assessment. Retrieved from: https://www.cambridgeassessment.org.uk/Images/111066-the-accuracy-of-forecast-grades-for-ocr-a-levels-.pdf

85 Gill, T. & Chang, Y. (2013). *The accuracy of forecast grades for OCR A levels in June 2012. Statistics Report Series No.64.* Cambridge Assessment. Retrieved from: https://www.cambridgeassessment.org.uk/Images/150215-the-accuracy-of-forecast-grades-for-ocr-a-levels-in-june-2012.pdf

86 Gill, T. & Chang, Y. (2015). *The accuracy of forecast grades for OCR GCSEs in June 2013. Statistics Report Series No.89.* Cambridge Assessment. Retrieved from: https://www.cambridgeassessment.org.uk/Images/241260-the-accuracy-of-forecast-grades-for-ocr-gcses-in-june-2013.pdf

87 Gill, T. & Benton, T. (2015). *The accuracy of forecast grades for OCR GCSEs in June 2014, Statistics Report Series No.91.* Cambridge Assessment. Retrieved from: https://www.cambridgeassessment.org.uk/Images/241265-the-accuracy-of-forecast-grades-for-ocr-gcses-in-june-2014.pdf

88 Gill, T. (2019). Methods used by teachers to predict final A Level grades for their students. *Research Matters: A Cambridge Assessment publication, 28,* 33-42. Retrieved from: https://www.cambridgeassessment.org.uk/Images/561974-methods-used-by-teachers-to-predict-final-a-level-grades-for-their-students.pdf

The extent to which an individual centre with a small entry may submit CAGs that are lenient or severe in a subject is unknown due to the weakness of the statistics. It is therefore not possible to confidently quantify the relative advantage/disadvantage an individual centre may have through greater reliance on the CAGs. Were evidence available to inform this calculation, it would be appropriate for it to be used for the determination of calculated grades.

While it is not possible to quantify and/or control for any relative advantage or disadvantage for centres where there is a greater reliance on the CAGs compared to those whose standardisation is performed fully with reference to the statistical model, centres with a larger entry can be protected from any absolute disadvantage. A significant limitation of **Option 1** is that, assuming a tendency towards leniency in the CAGs, the advantage potentially gained by students entering from centres with a small entry in a subject would lead to an absolute disadvantage to those students entering through larger centres. This is because those results from larger centres would be adjusted to be statistically severe in order to compensate for the leniency in outcomes for other centres; something that would not occur with **Option 2**.

The likely consequence of **Option 2** is that overall outcomes are likely to be lenient of the statistical predictions, to an extent that was unknown at the point the standardisation approach, including the handling of centres with a small entry, was being developed. Despite **Option 2** leading to almost inevitable leniency in the overall outcomes, it was decided to be the most appropriate course of action given the risks of absolute disadvantage to students who were part of a larger cohort from a centre.

8.4.3 Definition of a centre with a small entry

In order to define what constitutes a centre with a small entry, values of n_{thresh} and n_{small} must be set. There are three key factors that influence the setting of these thresholds which put upwards or downwards pressure on their positioning:

i. **Statistical appropriateness** – as discussed above, it would be inappropriate to standardise grades solely on a statistical basis where the cohort from a centre is particularly small for a given subject. This applies an upwards pressure on the thresholds.

ii. **Potential leniency** – given the likely optimism built into the CAGs, a greater reliance on this source of evidence is likely to lead to an overall inflation in the outcomes as discussed in 8.4.2. This applies a downwards pressure on the thresholds.

iii. **Consistency of treatment** – to ensure consistency of treatment, it is desirable to have as high a proportion of students as possible to be handled in the same way through the standardisation process. While this may vary from subject to subject, typically, centres with a small entry in a subject are likely to be in the minority. This applies a downwards pressure on the thresholds as this would increase the proportion of students treated in the same way.

In the absence of the entry data for 2020 at the time of performing the analysis to support the design of the approach, data from 2019 and the preceding years was used to model the potential effects of small centres. To reflect the CAGs, simulations were performed reflecting different levels of leniency. To present notable leniency in these CAGs, a generosity rate of 30% was selected based on high-end rate of over-prediction seen in the research literature.[84,85,86,87,88] This means that, at each grade for each centre, 30% of students are allocated to the grade above (except for the highest grade) based on the grade distribution actually achieved by each centre.

For context, the proportion of centres, and the number of students within those centres, falling below different $\bar{n}_{Hj}$ thresholds, based on the 2019 data, are shown in Tables 8.2 and 8.3, respectively.

Table 8.2. Breakdown of centre-level entries by qualification type

Qualification Type	Number of centres in 2019	Percentage of centres without historical data	Percentage of centres with…			
			$\bar{n}_H \leq 5$	$\bar{n}_H \leq 10$	$\bar{n}_H \leq 15$	$\bar{n}_H \leq 20$
GCSE (Phase 1)	15,019	4.2	5.1	10.0	13.6	16.5
GCSE (Phase 2)	49,071	9.4	5.8	12.9	21.7	30.4
A level	49,919	6.6	19.4	38.4	52.6	62.3

Table 8.3. Breakdown of student-level entries by qualification type

Qualification Type	Number of students in 2019	Percentage of students from centres without historical data	Percentage of students from centres with…			
			$\bar{n}_H \leq 5$	$\bar{n}_H \leq 10$	$\bar{n}_H \leq 15$	$\bar{n}_H \leq 20$
GCSE (Phase 1)	1,973,132	0.9	0.2	0.4	0.8	1.1
GCSE (Phase 2)	2,016,470	3.3	0.7	2.3	5.3	9.4
A level	733,015	2.0	3.0	10.0	18.5	26.6

It is clear from these summary data that, unsurprisingly, compared to GCSE, a higher proportion of students entered A level qualifications through centres whose current or previous entry size means they fall below the thresholds shown.

A full subject level breakdown is presented in Annex I. This is summarised for GCSE in Figure 8.6 at centre-level and in Figure 8.7 at student-level for threshold levels of 5 and 15. Equivalent plots are provided in Figures 8.8 and 8.9 for A level.

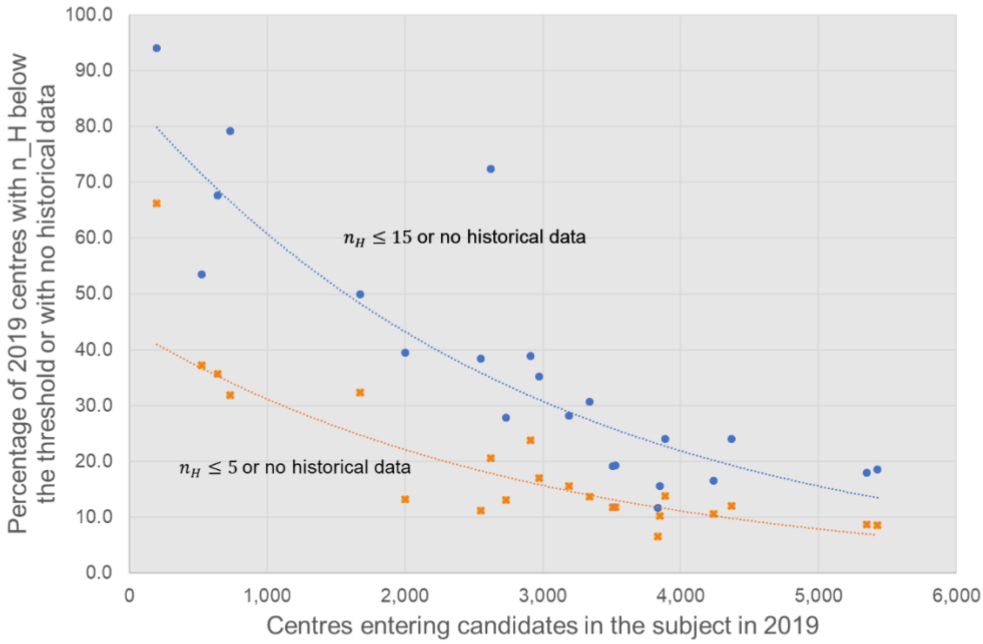


Figure 8.6. Relationship between the number of centres with entries in a GCSE subject and the percentage of those centres being considered as 'small' based on thresholds of 5 and 15.

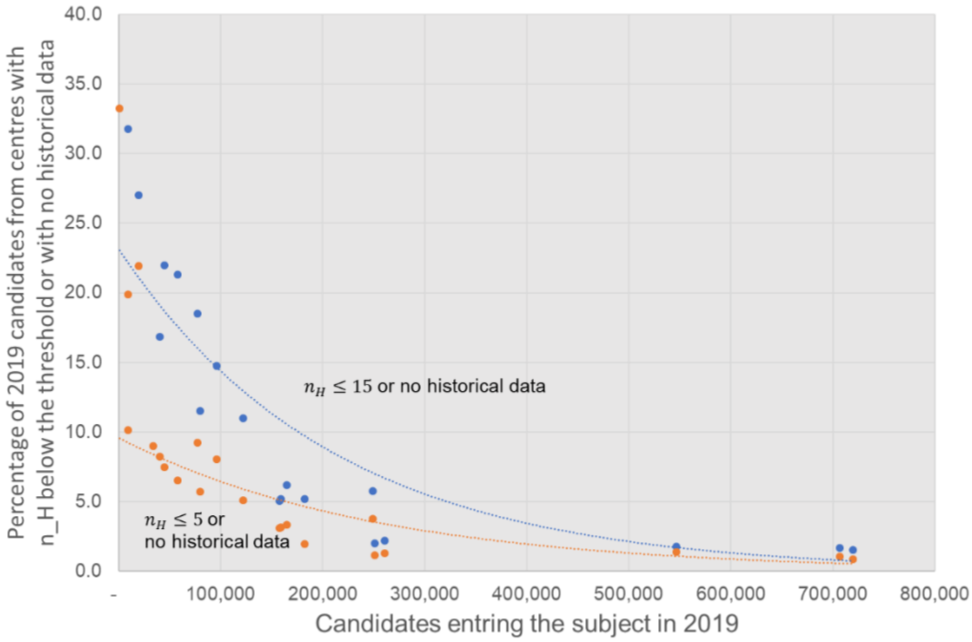


Figure 8.7. Relationship between the number of students entering a GCSE subject in 2019 and the percentage of those centres being handled as 'small' based on thresholds of 5 and 15.

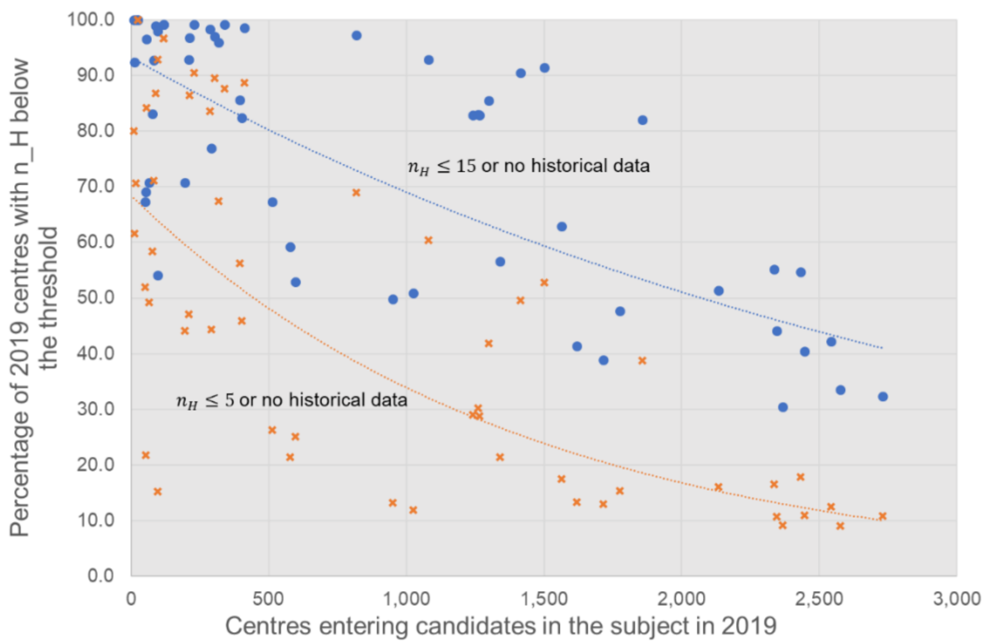


Figure 8.8. Relationship between the number of centres with entries in an A level subject and the percentage of those centres being considered as 'small' based on thresholds of 5 and 15.

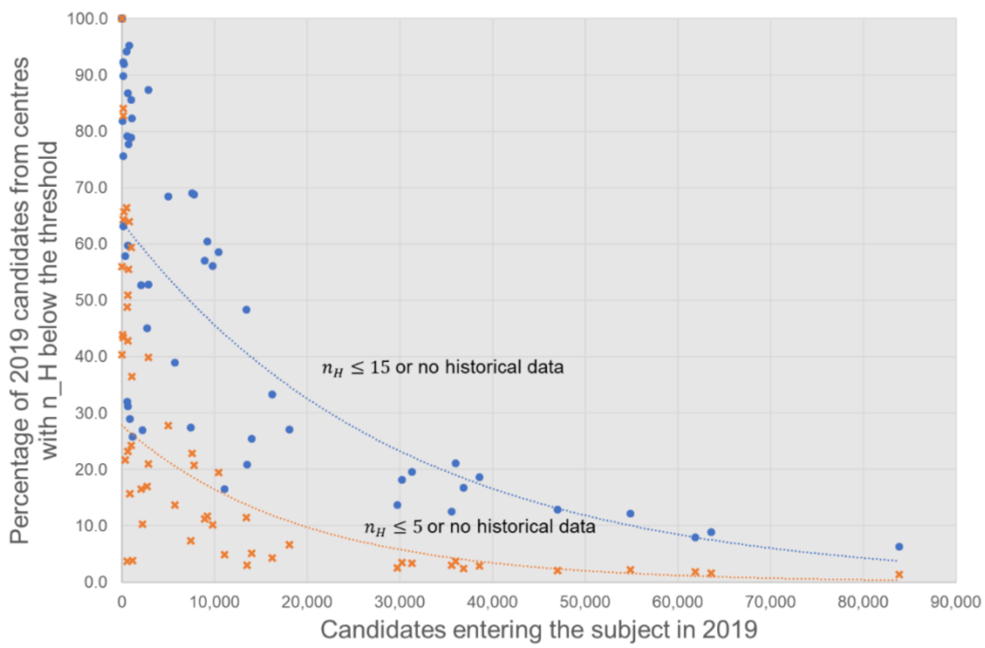


Figure 8.9. Relationship between the number of students entering an A level subject in 2019 and the percentage of those centres being handled as "small" based on thresholds of 5 and 15.

These data show that at GCSE, the significant proportion of students across subjects would be awarded grades based on the statistical model compared to those where greater emphasis is placed on the CAGs, based on the thresholds shown.

To identify the impact on outcomes for different configurations of the proposed approach, different combinations of n_{thresh} and n_{small} were simulated. For ease of interpretation and to provide an indication of sensitivity to different thresholds, the following three scenarios are presented here:

i. $n_{\text{thresh}} = 1$ and $n_{\text{small}} = 10$

ii. $n_{\text{thresh}} = 5$ and $n_{\text{small}} = 15$

iii. $n_{\text{thresh}} = 10$ and $n_{\text{small}} = 20$

Scenario i. is selected to represent the lowest possible configuration that could be potentially defensible. The overall qualification level leniency resulting from these different combinations of thresholds, based on the simulation using 2019 data, is presented in Table 8.4.

Table 8.4. Potential leniency at qualification level for different configurations of thresholds

		Deviation from prediction in cumulative percentage points								
		$n_{\text{thresh}} = 1$ and $n_{\text{small}} = 10$			$n_{\text{thresh}} = 5$ and $n_{\text{small}} = 15$			$n_{\text{thresh}} = 10$ and $n_{\text{small}} = 20$		
Qualification Type	**Number of students**	**Grade 7**	**Grade 4**	**Grade 1**	**Grade 7**	**Grade 4**	**Grade 1**	**Grade 7**	**Grade 4**	**Grade 1**
GCSE (Phase 1)	1,973,132	0.05	0.05	0.01	0.05	0.07	0.01	0.06	0.09	0.02
GCSE (Phase 2)	2,016,470	0.18	0.20	0.03	0.25	0.27	0.03	0.38	0.40	0.05
		A*	**A**	**E**	**A***	**A**	**E**	**A***	**A**	**E**
A level	733,015	0.29	0.43	0.08	0.57	0.89	0.15	0.92	1.51	0.23

It is clear from the figures presented above that the distribution of students across centres provides greater vulnerability to leniency at A level compared to GCSE for the same values of n_{thresh} and n_{small}. Despite these moderate levels of leniency, the levels of potential leniency vary significantly across subjects. A full breakdown of subjects based on this modelling is provided in Annex J with these data summarised in Figures 8.10 and 8.11 for GCSE grade 4 and A level grade A, respectively.

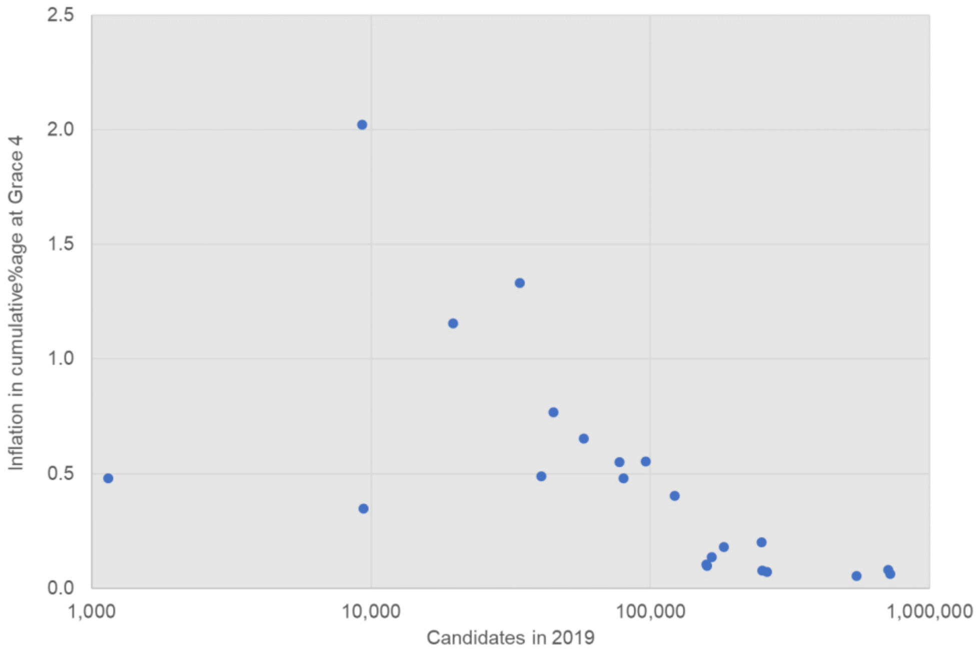


Figure 8.10. Relationship between the number of students entering a GCSE subject in 2019 and level of leniency at grade 4 based on a configuration of $n_{thresh} = 5$ *and* $n_{small} = 15$

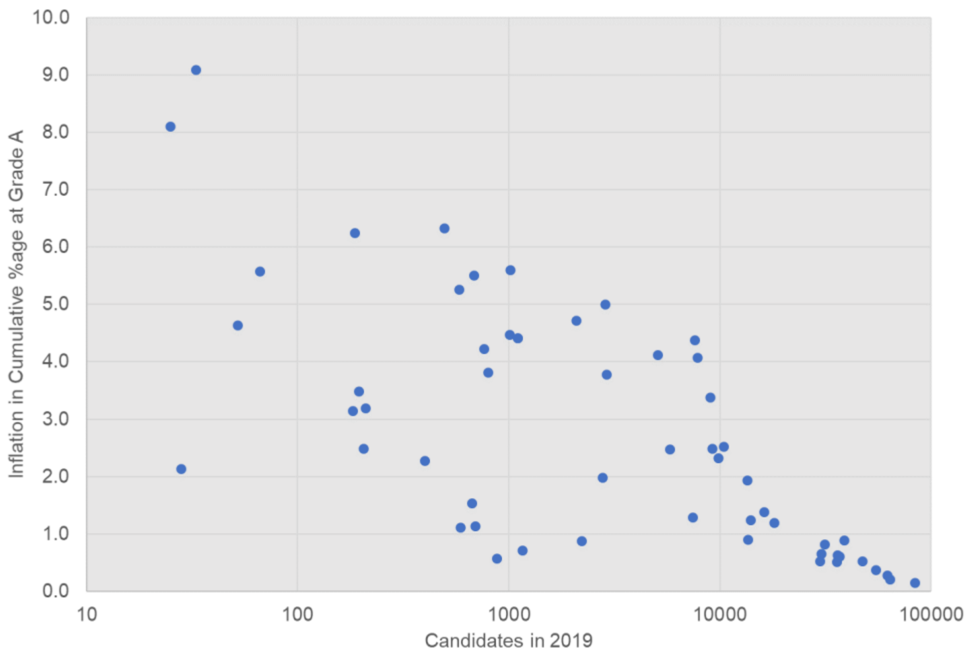


Figure 8.11. Relationship between the number of students entering an A level subject in 2019 and level of leniency at grade A based on a configuration of $n_{thresh} = 5$ *and* $n_{small} = 15$

Based on the 2019 modelling described, there were 14 A level subjects with levels of leniency greater than 5% at grade A* or A of which 3 had an entry over 1,000 students in summer 2019. This is provided in Table 8.5 for $n_{\text{thresh}} = 5$, $n_{\text{small}} = 15$ and $n_{\text{thresh}} = 1$, $n_{\text{small}} = 10$.

Table 8.5. A level subject with levels of inflation >5% (emboldened) based on modelled generosity. Figures are percentage points.

		$n_{\text{thresh}} = 5$, $n_{\text{small}} = 15$		$n_{\text{thresh}} = 1$, $n_{\text{small}} = 10$	
Subject	**Cand's in 2019**	**Leniency at A***	**Leniency at A**	**Leniency at A***	**Leniency at A**
Ancient History	681	2.6	**5.5**	1.9	4.6
Bengali	28	**9.6**	2.1	**8.8**	2.0
Biblical Hebrew	66	**9.1**	**5.6**	**6.4**	2.8
Classical Greek	210	**12.8**	3.2	**9.5**	2.4
Dutch	33	**5.5**	**9.1**	4.8	**8.6**
German	2,857	**5.9**	5.0	3.5	2.9
Gujarati	25	**5.3**	**8.1**	3.2	**6.6**
Italian	799	**8.7**	3.8	**6.7**	2.8
Japanese	195	**9.2**	3.5	**8.3**	3.0
Latin	1,103	**7.8**	4.4	4.7	2.7
Persian	205	**6.5**	2.5	**5.7**	2.2
Polish	1,004	**10.6**	4.5	**8.0**	3.2
Portuguese	494	**5.7**	**6.3**	4.7	**5.1**
Russian	695	**12.6**	1.1	**9.4**	0.9

From this list of subjects, only three had an entry of over 1,000 in 2019 with the most notable case being A level German. One option to manage potential leniency in these subjects due to the presence of centres with a small entry would be to set more stringent thresholds for either this subset of subjects or across all subjects. However, as can be seen in Table 8.4, even with the most stringent thresholds, that may arguably be tolerable, of $n_{\text{thresh}} = 1$ and $n_{\text{small}} = 10$ notable leniency would still be present in the outcomes.

Also, provided in Annex K, to build on these modelled analyses is a consideration of the statistical plausibility of performing statistical adjustments on particularly small numbers of students.

Based on discussion provided here and to strike the balance between statistical defensibility, the potential consequences for outcomes and commonality of handling across centres, the values of n_{thresh} and n_{small} were set to 5 and 15, respectively.

8.5 Standardisation of tiered subjects

The following tiered GCSE subjects are available in summer 2020:

Maths	Spanish	Panjabi
Statistics	Italian	Persian
Biology	Arabic	Polish
Chemistry	Chinese	Portuguese
Physics	Bengali	Russian
Combined	Gujarati	Turkish
French	Greek (Modern)	Urdu
German	Japanese	

The standardisation of tiered subjects this summer presents additional design challenges for the standardisation process to ensure fairness to students entering either tier and for those entering through centres which may have changed their tier entry strategy over recent years. The issues specific to the standardisation of tiered subjects are considered below.

8.5.1 Features of tiered subjects

In tiered subjects, students can either enter for the higher or foundation tier of the qualification. Tiered structures are put in place for qualifications which differentiate between students by task. Differentiation by task is where the attainment of students is typically defined by ability to correctly respond to a relatively large number of questions or assessment tasks with varying levels of difficulty. This contrasts with assessments which differentiate by outcome, where fewer, typically more open-ended, questions or tasks make up the assessments, with which students of all abilities can, ideally, engage. In these cases, the ability of students is determined by the quality of the response they provide rather than the difficulty of the questions themselves.

For qualifications that are intended for a wide ability range of students, such as GCSEs, there is the risk that assessments that differentiate by task may be poorly targeted across the range of abilities; for able students many of the questions may be too easy and/or, for less able students, many of the questions may be too difficult. In both cases, students' responses risk providing little or no assessment information which can be used to reliably differentiate between students. To mitigate these risks, tiered structures allow questions and tasks to be presented to students that are more appropriate to their capabilities.

Students entered for the higher tier are presented with more demanding assessment, typically, on broader more demanding subject content[89]. To reflect this increased demand, the grades available to students on the higher tier are grades 9 to 3[90] or grades 9-9 to 4-3 for GCSE combined science[91]. For foundation tier students, the grades available to them are grades 5 to 1 or grades 5-5 to 1-1 for combined science.

In a typical year, a relatively small proportion of students fail to register a performance sufficient to achieve the lowest grade available to those sitting the

89 The additional content required to be assessed on the higher tier of a qualification can been seen from reference to, for example, the DfE subject content for mathematics.

90 In tiered specifications, the grades available through the higher tier route are 9 to 4; however, to prevent students who narrowly miss a grade 4 from receiving an unclassified result, a safety grade – or "allowed grade 3" – is available. The grade 3 boundary is set half the number of marks below the grade 4 that grade 4 is below grade 5, so it is effectively a half grade, below which students are unclassified.

91 Students for combined science qualifications receive an award worth 2 GCSEs. The grade consists of two equal or adjacent grades from 9 to 1, giving 17 possible grade combinations, i.e. (9-9); (9-8); (8-8); through to (1-1). This is done so that students do not lose (or gain) two whole grades at each grade boundary. It ensures parity with students studying single sciences, where a student who misses grade 5 in biology by one mark would not also lose their grade 5 in physics or chemistry. On higher tier, the "allowed 3" is 4-3, which is equivalent to the half grade 3 on higher tier single awards.

higher tier[92]. In these cases, students are ungraded. The appropriateness of this approach, given the arrangements that are in place this summer needs to be considered along with the potential unintended consequences of whatever solution is in place.

Another key aspect of awarding tiered assessments relates to how the standards between the higher and foundation tiers are aligned. To provide supporting evidence to the alignment of tiers, the higher tier and foundation tier assessments share "common items". These are questions that are within the target ability range for both tiers based on content that is in scope for both tiers. The relative performance of students across the two tiers on these common items is then used to equate the tiers and set an appropriate inter-tier standard[93]. The absence of this assessment evidence this year, also needs to be factored into the arrangements.

A key design consideration for the standardisation of tiered subjects was, therefore, whether the process should operate at the tier-level (with each tier essentially being considered a separate subject) or at the overall subject-level (with no distinction being made between tier of entry). Inter-linked with this decision is how any issues of fairness between tiers would best be handled and whether the same restrictions in terms of the grades available would be appropriate, as outlined above.

8.5.2 Operation at the tier-level versus subject-level

The main advantage of operating the process at the individual tier level would be that the rank ordering of students within centres would be simplified. With the standardisation model operating at tier-level it would not be necessary for centres to rank order students across both tiers. The comparisons teachers would be required to draw would be between students that were preparing to be assessed on the same content and are more likely to have been taught together in the same groups. A logical consequence of this approach would also be to limit the grades available to students, in both the CAGs and calculated grades, to the tier to which they have been entered, as would be the case in a typical year.

There are, however, some key disadvantages of a tier-level approach. These primarily relate to the potential unintended consequences for centres changing their tier entry strategy between years and potential fairness to students across tiers.

The use of historical centre data for the purposes of standardisation, necessarily, makes a level of assumption about the stability of behaviour by that centre. It is known, however, that centres can change their tier entry strategy over time. If operating at the tier-level, this type of change may appear to suggest changes in the ability profile but that only exist at tier level rather than within the centre for the subject overall. These changes in tier entry strategy are more common early in the lifetime of specifications and may be particularly problematic in science and MFL subjects. This is because centres were encouraged to enter students for foundation tier to avoid the relatively high rates of students failing to secure a grade on higher tier[94] observed in the early years of these specifications.

[92] In 2019, 1.26% of students entered for the higher tier in GCSE mathematics, science, and modern foreign language titles received unclassified grades.

[93] Awarding and comparable outcomes

[94] GCSE tier entry in 2020

A change in the balance of entry across tiers could inadvertently advantage or disadvantage a centre overall and/or distort the relationship between tiers. An example is where a centre decides to enter students in 2020 for the higher tier who would typically have been their stronger students entered for the foundation tier. This would likely impact negatively on the centre's overall outcomes in the subject as illustrated in Figure 8.12.

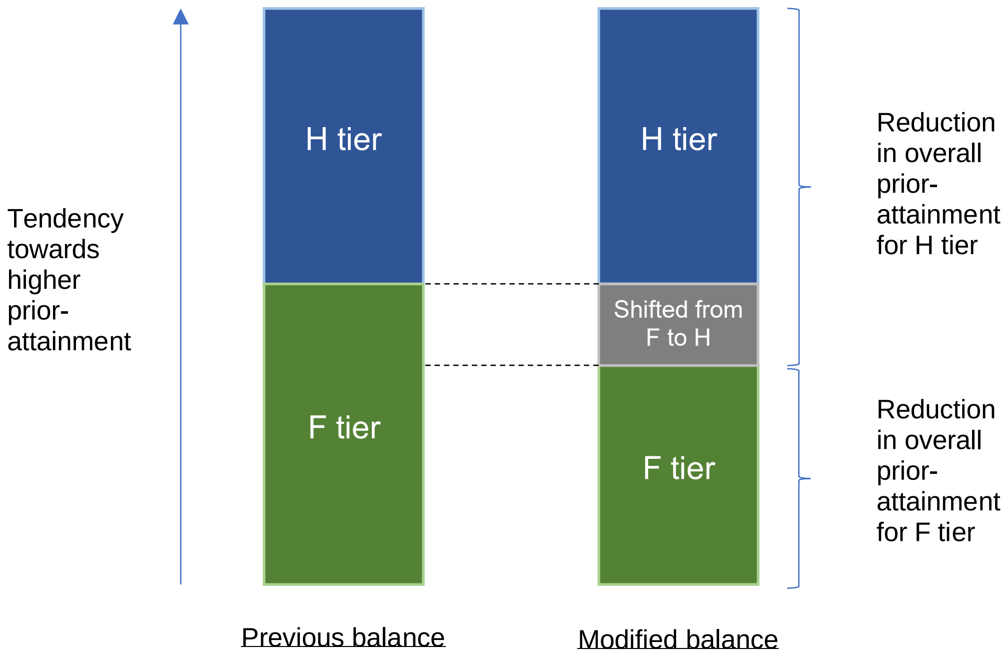


Figure 8.12. Reduction in mean prior-attainment for both higher and foundation tiers due to a change in entry strategy

This shows that, due to the change in entry strategy, the ability profile of both tiers would appear to have lowered while the ability profile for the centre overall would actually remain unchanged. This would likely lead to an inappropriate downward adjustment in prediction were the process to operate at the tier level.

Changes in entry strategy would also risk issues with the alignment of tiers leading to distorted rank orders within the subject. In the starkest cases, the standardisation process could lead to adjustments in different directions for each tier; one to address leniency and the other to address severity. This would mean that the rank order of students within centres would be disrupted or a complex approach attempting to reconcile adjustments across tiers for each individual centre would be required.

Operating the process at subject rather than tier-level would avoid the risks highlighted above. This approach does, however, require the collection of rank orders from centres across tiers, but would provide a solution to lack of technical information available to perform reliable linkage between tiers. Teachers' judgements of relative ability across-tiers is likely to be the most reliable source of evidence for equating this summer. It was noted above, however, that a key advantage of taking a tier-based approach was that comparisons between students would only be required between students who planned to sit the same tier. Given the likelihood of foundation and higher tier teaching groups being separated and the potential for them to have studied different content, a small number of teachers of tiered subjects were consulted over the feasibility of taking this approach. Having described the range of standardisation models under consideration at the time and having discussed the relative merits of taking different approaches and their experiences of changes to tier entry strategy, the consensus was that teachers ranking across tiers would be preferable.

8.5.3 Off-tier grades

As described in Section 8.5.1, one feature of the typical arrangements for tiered subjects is that the range of grades available to students sitting each tier are limited. The final aspect to be considered when deciding whether a subject-level approach to standardising tiered subjects was appropriate is whether students should continue to be limited to this range of grades or whether students could achieve 'off-tier' grades. This relates to both the CAGs that centres were permitted to submit and the final calculated grades awarded to students.

In relation to the CAGs, it is logical that grades should relate to the tier of entry to which students are entered. This was reflected in the guidance we published on 3 April[95] which included the following:

> *In the case of tiered GCSE subjects, schools and colleges should only provide centre assessment grades which reflect the tier of entry of the individual student (9 to 3 for higher tier; 5 to 1 for foundation tier).*

This statement reflects the overlap of the grading scale between tiers and was included to support teachers in making objective judgements regarding the grades students were likely to have achieved. Similar to tiering decisions in a typically year, an important consideration when allocating students to a tier of entry should be the grade they are likely to achieve[96]. It was therefore decided that centres would only be able to submit CAGs for tiered subjects on the tier to which the student has entered, with exam boards monitoring and, challenging where appropriate, any anomalous tier change requests made to existing entries or late entries with an atypical allocation across tiers.

The technical considerations regarding the potential award of off-tier calculated grades is, however, less straightforward. As discussed above, under normal circumstances, students failing to achieve the lowest grade available on the higher tier (grade 3 for single award and grade 4-3 for double award) would be ungraded. This is to reflect the failure of students to record a performance that is relevant to the grade range targeted by the assessment and to disincentivise inappropriate higher tier entries.

[95] Summer 2020 grades for GCSE, AS and A level, Extended Project Qualification and Advanced Extension Award in maths: guidance for teachers, students, parents and carers

[96] It is recognised that Grade 3 (and Grade 4-3 for double awards) on the higher tier are "allowed grades", included to provide protection against students under-performing in their assessments and failing to achieve the lowest grade that the higher tier assessments are design to target. It would, therefore, be unexpected, but not impossible, for a higher tier student to be allocated one of these grades as a CAG.

On the foundation tier, students are not able to achieve above a grade 5 (grade 5-5 for double awards[97]), irrespective of the level of performance they demonstrate. This is to reflect the demand of the assessment, the content coverage being different to that required for higher tier and to disincentivise inappropriate foundation tier entries. Consideration here relates to two scenarios:

> **Scenario 1)** Higher tier students at centres receiving a downward adjustment such that they would receive a calculated grade lower than the allowed grade range for the higher tier;
>
> **Scenario 2)** Foundation tier students at centres receiving an upward adjustment such that they would receive a calculated grade higher than those available for the foundation tier;

For **Scenario 1** to occur would require one of the following two situations:

i. a downward adjustment to be required for a centre with higher tier students who are allocated CAGs of grade 3 and for that adjustment to be of sufficient size that it cannot be achieved solely through the adjustment of foundation tier students who fall lower in the grade 3 rank order than all higher tier students. This is illustrated below.

Student A	Grade 3	H tier
Student B	Grade 3	F tier
Student C	Grade 3	H tier
Student D	Grade 3	F tier
Student E	Grade 3	F tier
Student E	Grade 3	F tier
Student G	Grade 3	F tier
Student H	Grade 2	F tier
Student I	Grade 2	F tier
Student J	Grade 2	F tier

Downward adjustment of these students is insufficient to realise the required adjustment

ii. a downward adjustment to be required for centres with higher tier students allocated CAGs of grade 4 that is so large it cannot be achieved solely through the adjustment of all students with CAGs of grade 3 and foundation tier students who fall lower in the grade 4 rank order than all higher tier students.

97 For simplification of discussion, only single award subjects are referenced from this point, however, the same principles are relevant for double awards.

Student A	Grade 4	F tier
Student B	Grade 4	H tier
Student C	Grade 4	F tier
Student D	Grade 4	F tier
Student E	Grade 4	F tier
Student E	Grade 4	F tier
Student G	Grade 3	F tier
Student H	Grade 3	F tier
Student I	Grade 3	F tier
Student J	Grade 3	F tier
Student K	Grade 3	F tier
Student L	Grade 3	F tier
Student M	Grade 2	F tier
Student N	Grade 2	F tier

Downward adjustment of these students is insufficient to realise the required adjustment

Scenario 2 would require an upward adjustment to be necessary for a centre of a size that could not be achieved solely through the adjustment of higher tier students who fall higher in the grade 5 rank order than all foundation tier students. This is illustrated below.

Student A	Grade 6	H tier
Student B	Grade 6	H tier
Student C	Grade 6	H tier
Student D	Grade 5	H tier
Student E	Grade 5	H tier
Student E	Grade 5	H tier
Student G	Grade 5	H tier
Student H	Grade 5	F tier
Student I	Grade 5	H tier
Student J	Grade 5	F tier

Upward adjustment of these students is insufficient to realise the required adjustment

When faced with these scenarios there are two options:

> **Option 1)** Prohibit the award of off-tier standardised grades, meaning higher tier students would be adjusted from grade 3 directly to ungraded and foundation tier students would be 'capped' at grade 5;
>
> **Option 2)** Allow the full grade range for standardised grades, irrespective of the tier of entry[98].

The advantage of **Option 1** is that it is consistent with the first aim of the standardisation process as outlined in Section 4.1: To award grades students were most likely to have achieved had they sat assessments this summer. It could be argued that **Option 2** is not consistent with this aim as they would result in the award

98 It is recognised that other options exist such as prohibiting the award of 'off-tier' grades for foundation tier students, but allowing the full range of grades to higher tier students or vice-versa. These options have been considered, but have been dismissed due to their asymmetric treatment of students across tiers and the potential behaviours that may be driven within centres to amend the tier to which their students are entered in light of this differential treatment.

of grades impossible to achieve on the tier of entry. However, prohibiting the award of off-tier grades in this way would have some notable drawbacks when it comes to potential fairness to students resulting from the practicalities of implementation. These issues are considered separately in the context of the two different scenarios.

Scenario 1 – downward adjusted higher tier students

When faced with scenario 1, **Option 1** would require one of the two following approaches to be taken:

> **Approach 1)** To stop the adjustment at the point the first higher tier student in the rank order would be moved to below a grade 3. This would leave the centre "under-adjusted".
>
> **Approach 2)** To "freeze" the grade for higher tier students at grade 3 who would otherwise have been adjusted to a lower grade, but allow the foundation tier students ranked higher than them to continue adjustment past them into grade 2.
>
> **Approach 3)** Allow higher tier students to 'fall off' the higher tier, meaning that those higher tier students who would otherwise have been adjusted to below a grade 3 are ungraded.

All of these approaches are problematic. Approach 1 would mean that the centre would be treated differently from others purely due to the location of a single higher tier student in their rank order leading to a different standard being applied.

Approach 2 would result in a modification of the rank order of students provided by centres, which is being taken as the most reliable measure of inter-student ability through this process.

Approach 3 would also result in a modification of the rank order of students provided by centres, but more substantially and with more significant negative implications for the affected students as they would fail to receive a grade at all based on this adjustment.

Scenario 2 – upward adjusted foundation tier students

When faced with scenario 2, **Option 1** would require one of the two following approaches to be taken:

> **Approach 1)** To stop the adjustment at the point the first foundation tier student in the rank order would be moved to above a grade 5. This would leave the centre "under-adjusted".
>
> **Approach 2)** To "freeze" the grade for foundation tier students at grade 5 who would otherwise have been adjusted to a higher grade, but allow the higher tier students ranked below them to continue adjustment past them into grade 6.

Approach 1 would be unfair to the higher tier students positioned below the first foundation tier student in the rank order, as their adjustment to higher grades would be "blocked" by the highest-ranking foundation tier student. Approach 2 would result in modification of the rank order submitted by the centre.

From the analysis presented here **Option 1** was deemed not to be tenable.

It is noted that allowing all grades to all students (**Option 2**) would have a tendency towards generosity; higher tier students would be protected from 'falling off' the tier and foundation tier students would be permitted a grade higher than grade 5. This generosity would be factored out through the standardisation process, meaning it is not a source of inflation. However, it would lead to a rank order of students that could not have been possible under the normal operation of exams. This is a consequence of prioritising the preservation of the centre rank orders, wherever possible.

The impact this might have on students around the grade 1/U borderline should be highlighted. It is likely that, through the process of standardisation, higher tier students would be failing to achieve a grade 3 at a lower rate than is usually the case. Also, if higher tier students are permitted to achieve off-tier grades to protect the rank order submitted by centres, they would be ranked above students receiving grade 1 on the foundation tier in a way that would not happen were they to have "fallen off" the higher tier. Despite the relatively small numbers of students that are usually in this position, they are proportionately large compared to the number of students being classified as ungraded. There were, therefore, concerns that this could lead to a notable change in the outcomes for students at the lower end of the foundation tier. This is because, due to the statistical approach to standard setting this year, grades 1 and 2 which would otherwise be awarded to foundation tier students, may be awarded to higher tier students leading to disadvantage. To address this issue, an adjustment was applied following an initial run of the process to protect outcomes for the foundation tier students. This is detailed in Annex L.

The final consideration relating to 'off-tier' calculated grades relates to the credibility of such an approach since there is a risk that foundation tier students may achieve grades usually awarded for performances on content they may not have studied. Through discussion with teachers of tiered subjects, it was noted that a common approach is to prepare students at the top of the foundation tier teaching sets for higher tier content even if the ultimate decision is taken to enter those students for the foundation tier. Given that these students at the top of the foundation tier are those most likely to achieve 'off-tier' grades above grade 5, this risk was deemed to be relatively low.

8.5.4 Summary of the approach for tiered subjects

Based on the analysis considered here, the following approach was selected:

- the standardisation process for tiered subjects would operate at the subject level,
- a single rank order of students across-tiers would be collected from centres,
- CAGs would be limited based on the tier of entry,
- calculated grades would not be limited based on the tier of entry,
- exam boards would monitor the number of students awarded 'off-tier' grades and address any potential disadvantage to students at the lower range of grades on both tiers arising from these arrangements

A discussion of inter-tier results, including the analysis of 'off-tier' grades will be provided in our follow-up report published on GCSE results day.

8.6 Standardisation of centres with a very large entry

The rank ordering of students in centres with a large entry in a subject represents a particular challenge to centres due to:

- the scale of the task;
- these centres often being FE colleges with a large number of resitting (and therefore closely grouped) students;
- the separation of teaching groups and teaching staff making the comparison of students across teaching groups particularly challenging.

On this basis, exam boards put in place arrangement for centres with large entries (>500 students) in GCSE English language and GCSE mathematics to group students in the rank order. This means that, rather than being individually ranked, students with CAGs between grades 4 and 1 were grouped into equal ranks in groups of 10. This grouping started at the top of the rank order for each grade, potentially leaving a group smaller than 10 students at the bottom of the rank order for that grade.

This was operationalised either by centres submitting rank orders as they would for any other subject, with each student assigned an individual rank with students within each group of 10 arbitrarily placed on the understanding that the exam board would apply the grouping post-hoc, or by centres submitting tied ranks for all students in a group.

The approach taken to the standardisation of students from these centres was that all students in a group of 10 (or smaller if at the bottom of the rank for a grade) would receive the same outcomes.

To implement the adjustment for these centres the following procedure was undertaken:

i. Include entries from very large centres in the standardisation process laid out in Section 8.2 considering students individually and disregarding any grouping that has taken place by the centre. Where exam boards have collected individual notional ranks for students these were used like any other rank for the purposes of this process. Where a single rank has been assigned to students, they were arbitrarily assigned a rank in the appropriate range to perform these initial stages of the process.

ii. Based on the CAGs provided at grades 4, 3, 2, and 1, the following procedure was performed for each centre at each grade:

    a. For the first group of ten students in the rank order at the current grade, identify the grades currently allocated to each student through step i.

    b. Allocate all students in that group the same grade as the highest graded students in that group.

    c. Repeat step ii. and b. for all groups at each grade.

This approach ensures that there is no disadvantage to students who have been rank ordered in this way and performing this stage in the process after the setting of cut-scores means that this benefit of the doubt does not have an adverse impact on the outcomes for students at any other centres.

8.6.1 Adjustment of Phase 4 GCSEs

Over recent years, GCSE, AS and A level qualifications have been revised through a phased programme of reforms. Summer 2020 marks the first award of the final phase – Phase 4 – of qualifications being reformed.

Subjects being awarded for the first time in their reformed versions are:

> A level: biblical Hebrew, Arabic, Bengali, Gujarati, Greek, modern Hebrew, Japanese, Panjabi, Persian, Polish, Portuguese, Turkish and Urdu
>
> GCSE: biblical Hebrew, Gujarati, Persian, Portuguese and Turkish

A requirement of the DCP approach, detailed in Section 8.2, is that historical data are available at each grade. For A level qualifications, this is not problematic since the grade scale used in the legacy specifications (A*-E) was retained through the reform process. This means that, as discussed in Section 7.2, it is possible to use centre performance in legacy versions of the qualifications for the purposes of standardisation this year. The situation with the Phase 4 GCSEs, however, is not as straightforward. Due to the change in grading scale between the legacy qualifications (A*-G) and the reformed versions (9-1), this information is not available in the historical data. An alternative approach was, therefore, required.

To provide a solution it is useful to draw on work that was conducted for potential use more broadly than just the Phase 4 GCSEs.

In a typical year, exam boards focus their attention on the *judgemental* grades[99] for the purposes of setting and maintaining standards. This enables sufficient statistical control and thorough scrutiny of students' work in a deliverable way to ensure that standards are maintained. A consequence of this approach is that the intermediate grade boundaries are set arithmetically meaning they are, notionally, evenly spaced.

The situation in summer 2020 is different given the absence of work to scrutinise and conventional grade boundaries to set. Using any of the approaches to statistical standardisation that were considered for use this summer, outcomes at every grade can be considered independently (and potentially adjusted independently) for each centre in each subject. Consideration was, however, given to whether the outcomes at only the key grades should be set statistically with the outcomes at the other grades (termed the arithmetic grades due to the approach taken to calculate the boundaries for these grades in a typical year) more closely following the distribution of CAGs submitted by the centre. This approach would have enabled a greater role for CAGs in the standardisation process. Due to the magnitude and the distribution of the optimism in the CAGs, this approach was not feasible. A full discussion of this issue is presented in Annex M.

[99] For the purposes of awarding, exam boards routinely focus both the evaluation of statistical evidence and the scrutiny of students' work at a few grades termed judgemental grades. At GCSE, these are Grades 7, 4 and 1. At AS and A level, these are grades A and E. Statistical monitoring is also performed at Grade 9 and A*. Collectively, these grades are referred to here as ***key*** grades.

In the case of Phase 4 GCSEs, however, this approach provides a necessary solution to the discontinuity in grade scales between the legacy and reformed versions of the qualifications.

At the point of transition from the legacy to the reformed versions of the qualifications, the statistical standard was carried forward mapping across particular grades[100]; having controlled for any changes in prior attainment, grade A on the legacy scale was statistically mapped across to grade 7 on the new scale, grade C was mapped to grade 4, and grade G was mapped to grade 1. While historical information is not available for each grade, it is available at these key points in the grade distribution. The standardisation of the Phase 4 GCSE subjects noted above was, therefore, performed based on the outcomes at these grades combined with the approach to standardisation at the arithmetic grades described in Annex M.

To perform this process, Steps 1 to 9, as detailed in Sections 8.2.1 to 8.2.9 were performed treating the qualifications as having only three available grade ranges: grades 9 to 7, 6 to 4 and 3 to 1 (plus ungraded)[101]. Once students have been allocated to these grade ranges a predicted grade distribution can be generated including the outcomes for arithmetic grades. This predicted distribution for the full range of grades is denoted as $P_j^* = \{P_{Mj}^* \ldots P_{kj}^* \ldots P_{0j}^*\}$, and was produced using the following procedure for each centre in each subject:

i. Calculate the cumulative percentage grade distribution, $\gamma_j = \{\gamma_{Mj} \ldots \gamma_{kj} \ldots \gamma_{0j}\}$, based on students' grades calculated in Step 9 (Section 8.2.9)

ii. Retain the cumulative percentage outcomes for grades 7, 4 and 1 as calculated in step i. by setting the predicted cumulative proportion outcomes, P_{kj}^*, to γ_{kj} for these grades.[102] This sets the outcomes at these grades in the same way as would be the case for any other GCSE qualification.

iii. To determine the level of adjustment that is to be applied at grade 9, the centre-level prediction for grade 9, P_{9j}^*, is produced by performing the following calculation:

$$P_{9j}^* = \left(1 - \frac{f_{7j} - \gamma_{7j}}{f_{7j}}\right) f_{9j}$$

where $f_j = \{f_{Mj} \ldots f_{kj} \ldots f_{0j}\}$ is the cumulative proportion grade distribution based on the CAGs submitted by centre j. Applying this equation has the effect of adjusting the outcomes based on the CAGs by proportionally the same amount as the adjustment applied at grade 7.

iv. For each arithmetic grade (grades 8, 6, 5, 3 and 2), calculate the revised predicted outcome, P_{kj}^*, as:

100 5 questions and concerns answered about new 9 to1 GCSE grading

101 For subjects with more than 500 prior-attainment matched students, a statistical prediction was available at grade 9 based on the calculation indicated on page 26 of our regulatory requirements in place for this summer: Requirements for the Calculation of Results in Summer 2020

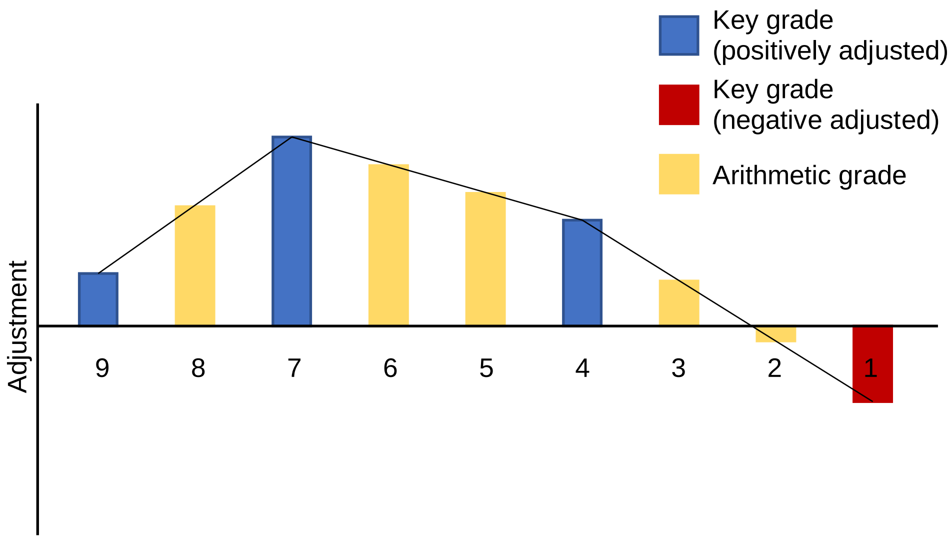
102 Note that for subjects with more than 500 prior-attainment matched students, grade 9 was included in this step and it was not necessary to perform step iii.

$$P^{*}_{kj} = \max\left(\gamma_{aj}, \min\left(\gamma_{bj}, f_{kj} + (\gamma_{bj} - f_{bj}) + ((\gamma_{aj} - f_{aj}) - (\gamma_{bj} - f_{bj}))\frac{(k-b)}{(a-b)}\right)\right)$$

where:

- k is the grade for which the revised predicted outcome is being calculated[103],
- a is the lowest key grade above the arithmetic grade,
- b is the highest key grade below the arithmetic grade,

This has the effect of applying a smooth adjustment at the arithmetic grades dependent on the adjustments at the key grades that surround it. This is shown graphically below:



v. Having formed this revised centre-level predicted grade distribution, the calculated grades are then formed by overlaying the centre's rank order onto their predicted cumulative percentage grade distribution, such that the proportion of students awarded each grade within the centre matches the predicted distribution as closely as possible.

8.7 Summary

This section has outlined the key stages in the approach taken to standardisation of CAGs in summer 2020. This process has been developed with a view to being as fair as possible across the widest range of students drawing on the most reliable sources of evidence available under different circumstances.

The next section provides a consideration of the CAGs that were used as an input to the process and the calculated grades that resulted from standardisation.

103 For adjustment purposes, the grades are converted to numerical values as follows: A level – A* = 6, A = 5…E = 1 and U = 0; GCSE – Grade 9 = 9, Grade 8 = 8…Grade 1 = 1, U = 0.

9 Summer 2020 results

This section provides an overview of the final results awarded to students in summer 2020, following the application of the standardisation approach outlined in Section 8. Overall results data is published by JCQ on results day each year, and to ensure consistency in our reporting we draw on those figures as far as possible. The data reported by JCQ includes overall results and outcomes by jurisdiction and subject. We focus on results for students in England, since we do not regulate reformed qualifications in Wales and Northern Ireland.

To supplement the data reported by JCQ, we have provided additional breakdowns of results. These analyses use data that was submitted to Ofqual by exam boards around a week before results were issued, and is similar to data that we collect every year. While this data is nearly complete (and indeed is likely to be more complete this year compared to other years),[104] there are likely to be some small differences in the overall number of students when compared to the published JCQ figures.

When considering the results for summer 2020 we focus on the calculated grades, since these are the final grades awarded to students. These grades are key, since they are the grades that students will use to progress and move on to the next stage of their lives. However, to provide the necessary background to the final results, we also consider the CAGs and the adjustments made to the CAGs as part of the standardisation process.

9.1 Context for summer 2020 results

To provide some context to the results for summer 2020, it is worth reflecting briefly on how the CAGs compare to outcomes in previous years, and how the distribution of centres with small cohorts differ across subjects. These factors are related and help to explain the results in particular subjects and overall. It is also worth considering how entries have changed and how that might impact on how results can be interpreted.

Tables 9.1 and 9.2 provide a summary of the CAGs for both AS and A level compared to outcomes in 2019 for students in England (cumulative percentages). These figures provide a useful indication of how the CAGs compare to previous outcomes. This is particularly the case at A level because although the cohort is slightly smaller in 2020 (partly due to a decline in the 18-year-old population but also due to some students not being able to receive a CAG or choosing not to for other reasons), the overall entry is large and reasonably stable. At AS, the cohort has continued to decline this summer. This means that making comparisons over time is less straightforward. We have therefore focused our discussion on A level, but where appropriate have provided corresponding figures for AS either in the text or an Annex.

The figures in Table 9.1 show that the A level outcomes based on the CAGs exceed outcomes in 2019 at all grades. There are clear differences across the grade range though, and the CAGs are more optimistic at grades A, B and C. This is likely to

[104] Any missing data is likely to be missing at random, so will not impact on the trends that we report.

reflect there being more students concentrated around these grades, and potentially these grades featuring more frequently in students' University offers.

Table 9.1. Centre assessment grades compared to 2019 outcomes – A level (cumulative %)

	Entry	A*	A	B	C	D	E
2019	736,734	7.7	25.2	51.1	75.5	90.8	97.5
CAGs	718,276	13.9	37.7	64.9	87.0	96.4	99.7
Difference	-18,458	6.2	12.5	13.8	11.5	5.6	2.2

Table 9.2. Centre assessment grades compared to 2019 outcomes – AS (cumulative %)

	Entry	A	B	C	D	E
2019	114,088	20.1	37.4	56.9	74.0	86.3
CAGs	70,505	26.0	48.6	72.2	87.8	96.6
Difference	-43,583	5.9	11.2	15.3	13.8	10.3

The optimism in the A level CAGs is likely a result of some centres believing that their students would have performed better this summer (some of whom would have been correct, and some of whom would have been incorrect), and the desire to give students the benefit of the doubt. The scale of the differences compared to 2019 are far beyond any normal variation in results that we typically see though. For example, in 2019, the cumulative percentage outcomes at grade A and above changed by 1%, and the pass rate by 0.1%. This means that, if the CAGs were awarded to all students this summer, the increase in overall outcomes would be significantly greater than observed in previous years. This would likely undermine the credibility of students' grades (see Section 5), and reinforces the need for standardisation (see Section 3).

Where possible, the standardisation process aims to ensure stability in overall outcomes. However, more reliance has to be placed on CAGs when awarding grades for centres with small cohorts. As outlined in Section 8, centres with small cohorts in a given subject either receive their CAGs or have greater weight placed on their CAGs than the statistical evidence. This is because it would be indefensible to statistically standardise when the number of students is very small. Given that the CAGs are generally optimistic, it is likely that final results will be higher than in previous years, particularly for those subjects with a higher proportion of centres with small cohorts.[105] This is a natural consequence of the model, the benefit of the doubt that has been included into its design, and the approach that has been taken for awarding grades to centres where the statistical evidence would be too unreliable.

Table 9.3 shows the proportion of centres with small cohorts for each A level subject with more than 500 matched entries (see Annex N for the full list of A level subjects). These figures reflect our previous analyses based on data from 2019 (see Section 8.4) and show some relatively large differences between subjects. This is not surprising given the uptake of different subjects nationally, and therefore the likely uptake within individual centres. However, it should be anticipated that, for those

105 Note that other decisions have been taken to give students the benefit of the doubt this summer that will also impact on results – for example, predictions were based on post-results grades and only those students that sat all of the assessments for a particular qualification in the reference year(s) (see Section 5).

subjects with a higher proportion of centres with small cohorts, outcomes will likely be higher than in previous years to a greater extent than for subjects that have a lower proportion of centres with small cohorts. This is a consequence of the standardisation model but is important to bear in mind when considering the overall outcomes and outcomes by subject below.

Table 9.3. Proportion of centres with small cohorts by A level subject (subjects with more than 500 matched entries)

Subject	Total number of centres	% of total centres		
		Small cohorts	Tapered adjustment	Statistical adjustment
Accounting	160	31.9	28.1	40.0
Ancient History	65	41.5	35.4	23.1
Art & Design: 3D Studies	147	51.0	29.9	19.0
Art & Design: Art, Craft and Design	801	38.1	47.3	14.6
Art & Design: Fine Art	1,671	28.4	51.0	20.5
Art & Design: Graphics	397	32.5	34.3	33.2
Art & Design: Photography	1,235	28.5	45.9	25.6
Art & Design: Textiles	483	56.5	32.5	11.0
Biology	2,555	8.1	25.8	66.0
Business Studies	1,740	11.2	29.6	59.2
Chemistry	2,509	11.6	30.8	57.5
Classical Civilisation	379	38.0	43.3	18.7
Computing	1,404	35.3	46.7	18.0
D&T: Fashion and Textiles	153	77.8	21.6	0.7
D&T: Product Design	1,255	51.2	44.7	4.1
Dance	214	50.0	41.1	8.9
Drama & Theatre Studies	1,309	30.0	54.2	15.7
Economics	1,657	13.3	26.5	60.2
English Language	984	13.7	34.8	51.5
English Language & Literature	563	19.0	39.6	41.4
English Literature	2,451	11.8	33.3	54.9
Environmental Studies	60	30.0	35.0	35.0
Film Studies	511	18.8	45.8	35.4
French	1,519	51.2	39.8	9.1
Further Mathematics	1,924	38.4	43.8	17.8
Geography	2,100	16.2	38.1	45.7
Geology	105	14.3	50.5	35.2
German	791	69.7	28.2	2.1
History	2,470	11.3	33.1	55.7
Latin	308	72.4	25.0	2.6
Law	536	16.4	28.0	55.6
Mathematics	2,730	9.1	20.0	70.9
Media Studies	993	10.2	37.3	52.6
Music	1,059	68.9	27.9	3.2
Music Technology	240	51.7	37.9	10.4
Philosophy	240	28.8	44.2	27.1
Physical Education	1,290	28.1	52.8	19.1
Physics	2,432	16.6	37.7	45.7
Politics	1,341	19.6	37.7	42.7
Psychology	2,403	8.3	20.4	71.2
Religious Studies	1,487	17.8	49.0	33.3
Sociology	1,715	10.6	24.5	64.8
Spanish	1,434	48.5	40.2	11.3
Statistics	55	30.9	21.8	47.3

The corresponding figures showing the proportion of centres with small cohorts for AS subjects are provided in Annex O. Given the declining AS entry, the relatively high proportion of centres with small cohorts in many subjects is not surprising. Nonetheless, similar to A level, there are differences between subjects. This suggests that any impact on outcomes is likely to differ by subject and will also be influenced by the extent to which the CAGs submitted by centres are optimistic. For AS, the extent of any leniency in the CAGs is less clear, because comparisons with previous years are less meaningful due to the large change in entries.

9.2 Overall outcomes in summer 2020

There are two ways in which we can consider overall outcomes. First, by considering the outcomes for matched students compared to predictions, and second, by comparing the outcomes for all students with previous years. The former involves making comparisons within the context of the awards made this summer – ie by comparing the statistical predictions with the outcomes for students matched to their prior attainment. This is how we evaluate exam boards' awards in a typical year because it is the most like-for-like comparison available – i.e. the predictions and outcomes largely include the same students, only those matched to their prior attainment. Although the approach to awarding grades is different this year, we have still used statistical predictions to guide the overall awarding process (see Sections 5 and 8). As such, it is useful to consider the matched outcomes compared to the predictions. And, because this is part of the awarding process, we consider this first. Note that the approach to standardisation (see Section 8) and the timing of when the national predictions were generated means that the number of students included in the predictions and matched outcomes differ slightly. This is due to the timing of when predictions are generated and is no different to a typical year.[106]

The second comparison considers the overall outcomes for all students compared to previous years. As in any year, the final outcomes for each qualification and subject[107] are published by JCQ.[108] These figures show the overall outcomes for all students and outcomes broken down by age group and jurisdiction. As in other years, we focus on the JCQ published figures for students in England. This is because we do not regulate reformed qualifications in Wales and Northern Ireland.

There are 2 points to note when making comparisons between overall outcomes this year and in previous years. First, the JCQ data is based on students' grades on results day (ie in August). These grades therefore do not include any changes that may occur due to post-results services such as reviews of marking and moderation. Such changes are typically in an upwards direction.[109] Second, because we are making comparisons across different years we are not necessarily comparing like-with-like if there have been changes in the cohort. Overall A level entries are reasonably stable, but there are likely to be differences at the individual subject level

106 Note that in a typical year some exam boards may re-generate their own predictions to only include those students that are due to receive a grade. This is not always the case though.

107 Note that subjects are grouped according to JCQ subject groupings.

108 JCQ: Examination results

109 Reviews of marking and moderation for GCSE, AS and A level: summer 2019 exam series

that could explain any differences in outcomes over time. At AS the entries have changed considerably, including in some individual subjects.

9.2.1 Outcomes for matched students compared to predictions

As in any year, we can review the outcomes of exam board’s awards by considering the extent to which the outcomes for students matched to their prior attainment align with the statistical predictions. Usually, we would do this separately for each exam boards’ specifications (where the entries are sufficiently large), by reviewing the differences between the outcomes and the predictions compared to a set of tolerances (that depend on the size of the entry). Given the approach to awarding grades this summer (see Section 5), our focus is on outcomes at the national level. Further, the usual reporting tolerances are not applied because of the way that the standardisation model operates (see Section 8), and because there is no evidence of student performance for senior examiners to review.

In a typical year, and for any given subject, the outcomes for students matched to their prior attainment do not generally align perfectly with the predictions. Grade boundaries are generally set to produce outcomes close to prediction (unless exam boards have evidence to do otherwise), but the precision of this depends on the distribution of students’ marks in the exam. As such, while outcomes are generally close to the predictions (unless exam boards have presented evidence to support an ‘out of tolerance’ award), it is not uncommon for outcomes in different specifications or subjects to differ from the predictions to a greater or lesser extent – and for some outcomes to be slightly above and others slightly below prediction.

The A level outcomes for matched students compared to prediction by subject and grade for summer 2020 are included in Table 9.4. Outcomes are presented for subjects where predictions were used as part of the standardisation process, i.e. those with more than 500 matched students. For almost all subjects (and at all grades), the outcomes are above prediction – and where outcomes are below prediction any differences are very minor. Given the approach this summer (see Sections 5 and 8), and that for A level there are many subjects with a relatively high proportion of centres with small cohorts (see Table 9.1), this is not surprising.

It is worth reflecting that there are differences in outcomes compared to prediction by subject though. When these differences are considered in the context of the CAGs and the proportion of centres with small cohorts, this is not surprising. Indeed, when we consider the outcomes relative to prediction for the subjects that exceed prediction at grade A by the greatest extent (see Table 9.4), unsurprisingly, these also tend to be the subjects that have a greater proportion of centres with small cohorts (see Table 9.1). These outcomes are therefore indicative of the standardisation approach operating as intended.

Table 9.4. A level matched outcomes relative to prediction by subject (subjects with more than 500 matched entries)

Subject	Matched students	Matched outcome relative to prediction					
		A*	A	B	C	D	E
Accounting	1,849	0.7	0.9	1.9	1.9	1.0	0.4
Ancient History	528	3.2	6.7	4.0	0.9	0.8	-0.1
Art & Design: 3D Studies	1,287	0.6	3.7	2.5	2.7	1.2	0.6
Art & Design: Art, Craft and Design	4,556	1.0	4.7	5.8	4.3	1.8	0.3
Art & Design: Fine Art	11,330	1.5	4.5	4.5	3.1	0.7	0.2
Art & Design: Graphics	4,138	0.2	1.4	2.2	1.6	0.6	0.2
Art & Design: Photography	9,701	0.7	2.7	4.2	3.2	1.2	0.3
Art & Design: Textiles	2,320	2.8	7.5	8.4	4.0	1.4	0.2
Biology	50,143	0.2	0.6	1.0	1.2	0.8	0.3
Business Studies	28,396	0.3	0.8	1.1	0.9	0.3	0.1
Chemistry	42,445	0.4	0.9	1.2	1.3	0.8	0.3
Classical Civilisation	2,247	3.7	7.9	5.6	2.3	0.3	0.1
Computing	10,303	3.8	6.7	8.1	7.9	4.6	1.4
D&T: Fashion and Textiles	595	7.9	19.1	22.5	17.9	7.0	2.0
D&T: Product Design	7,439	3.4	9.3	12.7	11.8	5.3	1.7
Dance	962	4.7	12.8	12.1	5.8	1.3	0.5
Drama & Theatre Studies	7,769	3.6	9.3	9.1	5.5	1.1	0.2
Economics	24,408	0.3	0.7	0.8	0.7	0.3	0.2
English Language	13,384	0.4	1.1	1.7	1.1	0.0	0.0
English Language & Literature	6366	0.4	1.6	2.6	1.7	0.3	0.0
English Literature	33,547	0.5	1.1	1.5	1.2	0.3	0.0
Environmental Studies	732	0.5	2.1	3.4	4.7	3.2	1.1
Film Studies	5,026	0.8	2.8	3.0	1.3	0.2	0.0
French	6,193	5.4	10.1	9.1	6.4	2.3	0.5
Further Mathematics	11,147	5.5	6.3	5.3	3.5	1.5	0.5
Geography	24,772	0.7	1.6	1.8	1.5	0.5	0.1
Geology	810	1.3	2.7	3.1	2.9	1.9	0.3
German	2,184	8.5	15.8	15.0	9.6	3.5	0.6
History	36,214	0.5	1.3	1.5	1.0	0.3	0.1
Latin	738	19.2	19.7	11.2	4.5	1.5	0.6
Law	9,868	0.2	0.8	1.5	1.3	0.7	0.3
Mathematics	59,752	0.2	0.4	0.7	0.7	0.5	0.2
Media Studies	12,218	0.2	1.1	1.2	0.7	0.2	0.0
Music	3,298	7.6	17.5	19.5	12.0	4.0	0.7
Music Technology	1,046	2.8	8.8	13.7	13.0	5.7	1.5
Philosophy	1,959	3.1	5.5	6.3	4.3	2.5	0.4
Physical Education	9,452	2.1	5.7	7.0	5.9	3.0	0.9
Physics	29,049	0.9	1.7	2.3	2.5	1.7	0.7
Politics	13,428	1.0	2.0	2.3	1.7	0.7	0.3
Psychology	55,115	0.2	0.5	0.7	0.6	0.3	0.1
Religious Studies	12,803	1.5	3.1	3.5	2.6	0.9	0.2
Sociology	32,308	0.2	0.6	0.6	0.6	0.2	0.1
Spanish	6,522	5.9	10.3	9.9	6.2	2.0	0.4
Statistics	851	0.3	1.3	0.9	0.6	0.5	-0.1

The matched outcomes compared to predictions for AS are provided in Annex P. A similar effect is observed to A level, with differences in the extent to which individual subjects exceed the statistical predictions. Given the relatively low number of entries for AS though, the number of subjects using statistical predictions is low. However, as for A level, those subjects that exceed predictions by the greatest extent tend to be those with a higher proportion of centres with small cohorts (see Annex O.

9.2.2 Outcomes for all students compared to 2019

A second way to consider outcomes is to compare the outcomes for all students to the corresponding figures from 2019, at overall qualification level (AS or A level) or by subject. While we generally expect results to be reasonably stable from one year to the next, there can be small fluctuations in outcomes due to normal variation. There can also be fluctuations due to changes in entry and the type of students entering particular qualifications. This could occur at the overall cohort level (eg if there was a significant shift in the age profile of entrants), but is far more likely at the individual subject level where the type of students entering a particular subject might change.[110] For example, if there was a shift of stronger students into a particular A level subject, then outcomes in that subject would likely increase. Such changes are possible this year, as in any year. We also know that this year some students were not able to receive a CAG or may have chosen to delay entering until a subsequent exam series. These changes may also impact on the extent to which outcomes differ from previous years.

The outcomes for all students in England compared to 2019 are published by JCQ.[111] Outcomes are grouped according to the subject categories used by JCQ, meaning that some subject groups – e.g. other modern foreign languages – include multiple subjects. As such, the JCQ subject groups do not align precisely with the subject groupings that have been used for running the standardisation process (and therefore the groupings for which statistical predictions were generated and outcomes reported).[112]

Overall, A level results in England have increased by 2.4% at grade A and above compared to 2019. This is a larger change than observed in a typical year (eg there was a 1% decrease in outcomes between 2018 and 2019), and is likely to reflect the approach to awarding grades this summer, as outlined in Sections 5 and 8. This includes the approach to awarding grades for centres with small cohorts, and the desire to give students the benefit of the doubt where possible as part of the standardisation process. The entries for A level compared to 2019 are reasonably stable, even in the context of circumstances this summer, meaning that it is possible to make such comparisons over time.

While overall results have increased compared to 2019, the outcomes in individual subjects differ to varying degrees relative to 2019. For ease of reference we have re-produced the JCQ figures showing the outcomes by subject compared to 2019 (Table 9.5). It should not be surprising that the subject groups with the largest increase in outcomes compared to 2019 tend to be those that exceeded the

110 Note that these sort of changes would not be evident in the matched outcomes compared to prediction, since they are comparing a similar group of students within the same year, rather than over time.
111 JCQ: Examination results
112 See Annex A of our Requirements for the calculation of results in summer 2020

predictions to the greatest extent, and are those with a larger proportion of centres with small cohorts.[113] Again, this is indicative of the modelling approach operating as intended.

Table 9.5. A level subjects with greatest increase in outcomes at grade A and above in 2020, compared to 2019 (JCQ subject groupings)[114]

Subject	Entry	Differences for all students compared to 2019					
		A*	A	B	C	D	E
Art and design	38,916	0.8	3.5	2.9	1.6	0.3	0.1
Biology	59,473	0.7	1.2	2.0	2.7	2.1	1.1
Business studies	32,794	0.2	1.1	1.6	1.9	1.3	0.8
Chemistry	51,850	1.0	2.0	2.5	3.2	2.8	1.2
Classical subjects	4,599	7.7	10.4	7.9	4.0	1.2	0.3
Computing	11,607	4.8	9.3	9.7	9.5	6.0	2.6
Design & technology	9,167	4.1	10.8	13.7	12.7	5.7	1.9
Drama	8,667	4.1	9.9	10.3	6.2	1.9	0.4
Economics	30,367	0.7	1.5	2.1	2.4	1.2	0.6
English language	14,666	0.1	0.1	-2.7	-2.0	-0.6	0.2
English literature	37,964	0.6	1.3	2.1	2.2	1.0	0.3
English language & literature	7,156	0.2	1.7	2.3	1.9	0.3	0.3
French	7,557	5.5	9.6	5.4	4.4	2.1	0.7
Geography	27,309	0.6	1.7	2.4	2.8	1.2	0.5
German	2,663	8.3	13.3	12.4	7.0	2.4	0.4
History	40,849	1.0	1.6	2.3	2.4	1.3	0.4
Law	11,090	0.1	0.7	1.6	1.4	0.6	0.5
Maths	87,164	0.3	0.8	0.7	1.0	0.9	0.7
Maths (further)	14,125	7.9	8.7	7.1	5.1	2.6	1.1
Media/Film/TV Studies	19,517	0.6	2.8	4.5	3.1	0.9	0.4
Music	5,031	7.1	16.5	18.9	12.5	4.4	1.2
Performing/Expressive Arts	1,060	4.5	12.9	14.0	6.8	1.8	0.7
Physical Education	10,250	2.2	5.8	7.0	6.5	4.0	1.3
Physics	34,996	1.7	3.2	1.7	1.7	1.9	1.5
Political studies	16,220	1.2	2.2	2.0	1.6	1.0	0.5
Psychology	62,544	0.1	0.5	0.9	1.9	1.8	0.4
Religious studies	14,567	1.7	3.7	4.7	4.0	1.7	0.4
Sociology	36,789	-0.1	0.2	1.8	1.5	0.6	0.4
Spanish	8,033	5.9	8.9	7.3	4.5	1.3	0.5
Other sciences	1,475	3.4	5.6	6.3	5.1	2.8	0.2
Other subjects	5,392	1.8	3.8	5.2	5.3	3.5	1.4
Other MFL	4,599	8.4	14.6	11.1	6.9	3.2	1.3

Given the falling entry at AS,[115] it is difficult to make meaningful comparisons over time for individual subjects. The JCQ published figures show that, while overall outcomes are relatively stable (increasing by 1% this summer compared to 2019), there are significant shifts in particular subjects that are likely to reflect both the

113 Note that Table 9.5 reports outcomes by JCQ subject group, while those in Tables 9.3 and 9.4 report at the individual subject level.
114 Retrieved from https://www.jcq.org.uk/examination-results/
115 Provisional entries for GCSE, AS and A level: summer 2020 exam series

changing entry and the approach to awarding grades (including the impact of small cohorts).

9.3 CAGs compared to final grades

The overall results for A level have increased compared to 2019, but they are lower than the outcomes shown in Section 9.1 based on the CAGs. This is inevitable given that the CAGs were generally optimistic and the ministerial direction[116] that, as far as possible, overall results should be similar to previous years (see Section 2). The same is observed at AS, but given the large change in entry, it is more difficult to interpret these differences.

Tables 9.6 and 9.7 show the outcomes for all students in England based on the CAGs and the final grades awarded to students. For these figures we have used the data submitted to us by exam boards to ensure that we are comparing like-with-like (ie we have included the same students in both calculations). For A level the differences are greatest at grades A, B and C. This is likely to reflect that the CAGs tended to be most optimistic at those grades.

Table 9.6. Centre assessment grades compared to 2020 outcomes – A level

	Entry	A*	A	B	C	D	E
CAGs	718,276	13.9	37.7	64.9	87.0	96.4	99.7
2020	718,276	8.9	27.6	53.8	78.0	92.3	98.2
Difference	0	-5.0	-10.1	-11.1	-9.0	-4.1	-1.5

Table 9.7. Centre assessment grades compared to 2020 outcomes – AS

	Entry	A	B	C	D	E
CAGs	70,505	26.0	48.6	72.2	87.8	96.6
2020	70,505	21.1	39.4	60.5	77.6	89.7
Difference	0	-4.9	-9.2	-11.7	-10.2	-6.9

To provide a more detailed breakdown, Table 9.8 shows the percentage of CAGs that were adjusted or unadjusted for AS and A level. For both AS and A level the majority of CAGs were unadjusted (58.7% and 59.8%, respectively). This means that most students received the same grade as the grade submitted by the centre. Where there was an adjustment, this was usually by 1 grade. Overall, 96.4% of A level and 91.5% of AS grades awarded were the same as, or within 1 grade, of that submitted by the centre.

116 Direction issued to the Chief Regulator of Ofqual, 3 April 2020

Table 9.8. Percentage of CAGs unadjusted and adjusted[117]

	Total	Adjusted down by			Unadjusted	Adjusted up by		
		3+ grades	2 grades	1 grade		1 grade	2 grades	3+ grades
A level	718,276	0.2	3.3	35.6	58.7	2.2	<0.1	<0.1
AS	70,505	1.1	7.1	28.7	59.8	2.9	0.2	<0.1

9.4 Regional outcomes

To provide a further breakdown of results, we have included data on the cumulative percentage outcomes at grade A and above by region in England, compared to previous years. This data is published by JCQ each year and outcomes for all grades are available via the JCQ website. Table 9.9 shows that the overall increase in outcomes at A level (+2.4%) is reflected in the regional data. Outcomes have increased in all regions, and while these increases differ slightly in scale, they reflect normal variation in a typical year. For example, in 2019 the regional outcomes changed by between +0.2% and -2.1% compared to the previous year, while the differences this year range from +1.8% to +3.4%. There is no evidence to suggest that the process of awarding grades this year has had a differential effect across regions.

Table 9.9. Regional outcomes at grade A and above (2018 – 2020)[118] (percentage)

Region	2018	2019	2020	Change 2018 to 2019	Change 2019 to 2020
North East	22.8	23.0	24.9	0.2	1.9
North West	24.6	23.5	25.3	-1.1	1.8
Yorkshire and the Humber	24.4	23.2	25.4	-1.2	2.2
West Midlands	22.6	22.0	24.2	-0.6	2.2
East Midlands	23.1	21.0	24.4	-2.1	3.4
Eastern Region	26.4	25.6	28.0	-0.8	2.4
South West	26.8	25.8	28.6	-1.0	2.8
South East	29.3	28.3	30.7	-1.0	2.4
London	28.1	26.9	29.8	-1.2	2.9

9.5 Outcomes by centre type

We have also considered overall outcomes by different types of centre compared to previous years. We have grouped centres to reflect the level of the qualifications included in this report. As such, 'other' centres include secondary modern schools, free schools and institutions such as tutorial colleges, language schools, special schools, pupil referral units (PRU) and training centres. The figures that are

117 Percentages rounded to 1 decimal place so may not sum to 100.
118 Retrieved from https://www.jcq.org.uk/examination-results/. Differences calculated on rounded figures.

presented are based on data submitted to Ofqual by exam boards around a week before results were issued.

In considering outcomes by centre type we have focused on A level because the entries for AS are small and declining, meaning that it is difficult to make comparisons over time. Further, when breaking down the outcomes by centre type, the number of centres in each group is relatively low.

Tables 9.10 and 9.11 provide the A level outcomes at grade A and above and grade C and above, respectively, by centre type from 2018 to 2020. These figures only include centres that had entries in all 3 years. While this facilitates making comparisons over time, it does not necessarily follow that the number of students will have remained stable within individual centres. As such, changes in outcomes can still reflect a changing cohort.

Overall, outcomes have increased relative to 2019 for all centre types at both grades. The extent of the increase varies by centre type though. This reflects differences in a typical year. Indeed, in 2019, the grade A outcomes for some centre types increased relative to the previous year, while others decreased.

Table 9.10 Outcomes by centre type at grade A and above (2018 – 2020) (percentage)

Centre type	Total centres	2018	2019	2020	Change 2018 to 2019	Change 2019 to 2020
Secondary comprehensive	510	21.3	19.8	21.8	-1.5	2.0
Secondary selective	72	37.5	36.0	37.2	-1.5	1.2
Independent	530	45.8	43.9	48.6	-1.9	4.7
Sixth form/FE/tertiary	219	21.6	20.5	20.8	-1.1	0.3
Academy	1,162	24.9	23.6	25.3	-1.3	1.7
Other	150	23.6	24.4	28.0	0.8	3.6

Table 9.11 Outcomes by centre type grade C and above (2018 – 2020) (percentage)

Centre type	Total centres	2018	2019	2020	Change 2018 to 2019	Change 2019 to 2020
Secondary comprehensive	510	74.4	72.1	74.6	-2.3	2.5
Secondary selective	72	84.3	83.4	83.6	-0.9	0.2
Independent	530	88.6	87.6	89.9	-1.0	2.3
Sixth form/FE/tertiary	219	75.2	73.4	73.7	-1.8	0.3
Academy	1,162	76.9	75.2	77.1	-1.7	1.9
Other	150	69.8	69.5	74.5	-0.3	5.0

9.6 Grade combinations

The results that we have reported thus far have considered overall outcomes, but we have also analysed results for individual students and in specific subjects (focusing on A level maths, the largest entry subject – see Section 9.8). Our student level analyses focus on grade combinations for students that entered 3 A levels in summer 2020, since this is the majority of A level students. This facilitates making comparisons over time.[119]

Figure 9.1 shows the grade profiles for students taking 3 A levels (ie the grades they achieved across the 3 qualifications), with attainment decreasing from left to right (3 A* grades on the left of the plot and 3 U grade on the right). Because there are a large number of different grade combinations possible when a student has sat 3 A levels, we have combined some groups by using the # symbol to represent any grade that is lower than the first grade in each combination, eg CC# represents students awarded 2 C grades and a grade D, E, or U. The solid lines on the graph show the trajectories for 2017 to 2019, and the points show the results based on the summer 2020 grades. The percentages are cumulative such that they include the percentage of students achieving each grade combination or a higher combination. For example, the cumulative percentage for AAA will also include students achieving A*AA, A*A*A, and A*A*A*. For reference, the underlying data showing the percentages and cumulative percentages for each grade combination are provided in Table 9.12.

Figure 9.1 shows that, in general, the grade combinations for individual students this summer are distributed very similarly to previous years. This can be seen clearly when the cumulative percentage of students awarded sets of grade combinations is visualised, with the grade combinations in 2020 closely tracking the results in previous years. This suggests that, for students sitting 3 A levels, the grade combinations awarded are generally similar to previous years. This is important when considered in the context of admissions to HE, since this relies on there being general stability within the system from one year to the next (see Section 5). As such, if there was no standardisation process this summer and students were awarded the CAGs, the grade profiles would likely have looked very different (see Section 9.1). This would have likely had implications for HE admissions.

119 The figures for previous years are based on final results awarded to students (ie they include any post-results changes).

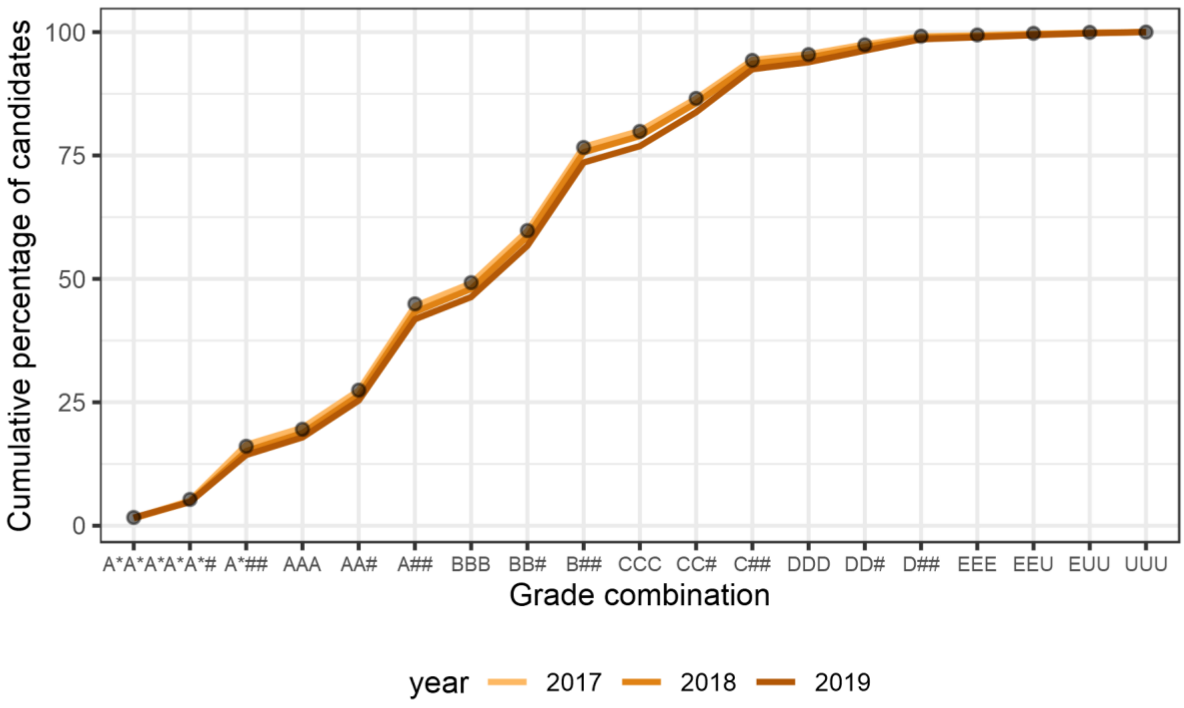


Figure 9.1. Cumulative percentage of students by grade combinations

Table 9.12. Percentage and cumulative percentage of total students awarded grade combinations between 2017 and 2020

	Percentage of total students				Cumulative percentage of total students			
Grade combination	2017	2018	2019	2020	2017	2018	2019	2020
A*A*A*	1.5	1.5	1.6	1.7	1.5	1.5	1.6	1.7
A*A*#	3.6	3.4	3.1	3.7	5.1	4.8	4.8	5.3
A*##	11.3	10.3	9.6	10.8	16.4	15.1	14.3	16.1
AAA	3.5	3.7	3.6	3.5	19.8	18.8	17.9	19.5
AA#	7.5	7.6	7.5	7.9	27.4	26.4	25.4	27.5
A##	17.1	16.9	16.5	17.5	44.5	43.3	41.8	44.9
BBB	4.6	4.7	4.5	4.3	49.1	48.0	46.3	49.2
BB#	10.6	10.6	10.4	10.6	59.7	58.6	56.7	59.8
B##	17.0	17.1	16.9	16.8	76.7	75.6	73.6	76.6
CCC	3.2	3.3	3.3	3.3	80.0	78.9	76.9	79.9
CC#	6.5	6.7	6.9	6.7	86.5	85.6	83.8	86.6
C##	7.8	8.0	8.7	7.7	94.3	93.6	92.5	94.3
DDD	1.2	1.2	1.4	1.2	95.5	94.8	93.9	95.4
DD#	2.0	2.1	2.3	2.0	97.5	97.0	96.2	97.4
D##	1.6	1.9	2.3	1.8	99.1	98.9	98.6	99.2
EEE	0.2	0.3	0.4	0.3	99.4	99.2	98.9	99.4
EEU	0.3	0.4	0.4	0.3	99.6	99.5	99.4	99.7
EUU	0.2	0.3	0.4	0.2	99.9	99.8	99.8	99.9
UUU	0.1	0.2	0.2	0.1	100.0	100.0	100.0	100.0

To provide further detail on specific grade combinations, Table 9.13 shows the percentage of students awarded the top 20 grade combinations over time. This shows that any variation this summer is similar to differences observed in previous years. For example, the percentage of students awarded 3 grade As this summer differs by only 0.1% from 2019, and the percentage of students awarded 3 grade Cs is the same as in 2019. Again, this suggests stability over time in terms of the grade combinations awarded to students.

Table 9.13. Proportion of students achieving the 20 most common grade combinations (2017 to 2020)

	Percentage of total students			
Grade combination	2017	2018	2019	2020
BBC	7.9	7.9	7.8	7.7
BCC	7.0	7.1	7.1	6.9
ABB	7.0	7.0	7.0	6.9
AAB	5.9	6.0	5.9	6.0
BCD	5.4	5.4	5.3	5.4
CCD	4.9	5.0	5.2	4.9
ABC	5.2	5.1	4.9	5.3
BBB	4.6	4.7	4.5	4.3
A*AA	4.2	4.1	3.8	4.1
CDD	3.6	3.7	4.0	3.5
AAA	3.5	3.7	3.6	3.5
CCC	3.2	3.3	3.3	3.3
A*AB	3.2	2.8	2.7	2.9
CDE	2.6	2.5	2.9	2.5
A*A*A	2.8	2.7	2.5	2.9
BBD	2.2	2.2	2.1	2.2
DDE	1.6	1.7	1.9	1.6
ACC	1.6	1.6	1.6	1.7
A*A*A*	1.5	1.5	1.6	1.7
BDD	1.6	1.6	1.5	1.5

9.7 Grades in maths and further maths

A level maths is the largest entry A level, and a proportion of students that sit this qualification also sit A level further maths. While most students sit both A level maths and further maths in Year 13 (typically aged 18), a minority of students sit maths in Year 12 and further maths in Year 13.

To consider the relationship between students' grades in maths and further maths we have compared the outcomes in these subjects for students entering both qualifications in 2020, with students entering both qualifications in 2019.[120] This allows us to consider the extent to which students achieved the same or different grade in each subject, in comparison to previous years. Because of the potential impact of centres with small cohorts on outcomes (see Section 9.1), we do not necessarily expect the relationship this summer to precisely replicate that observed

120 Note that we have only included students that entered both qualifications this summer and therefore have a calculated grade in each.

in previous years. This is particularly the case given that the proportion of centres with small cohorts is higher in further maths.

Table 9.14 provides a summary of students' grades in maths compared to further maths for students entering both qualifications in 2019 or 2020. While the majority of students received a higher grade in maths than further maths, the percentage of such students is lower this summer (57.4%) compared to 2019 (73.3%). Given that there are a higher proportion of centres with small cohorts in further maths – and that this is related to more lenient outcomes – this is not surprising (see Section 9.1). It also follows that in 2020, a greater percentage of students received the same grade in both qualifications (39.5% in 2020 compared to 26.5% in 2019), or a higher grade in further maths than maths (3.1% in 2020 compared to 0.2% in 2019). The proportion of students receiving a higher grade in further maths is still very low though.

Table 9.14. Outcomes in maths and further maths grades in 2019 and 2020 (%)

	Higher grade in maths	Same grade	Higher grade in further maths
2019	73.3	26.5	0.2
2020	57.4	39.5	3.1

9.8 Summary

In summary, overall A level results in England have increased at all grades this summer compared to 2019. These increases are generally greater than the changes observed in a typical year, and reflect the approach to awarding grades this summer. This includes the approach to awarding grades for centres with small cohorts, and the desire to give students the benefit of the doubt where possible as part of the standardisation process. While overall results have increased, they are lower than the outcomes based on the CAGs. This is inevitable given that the CAGs were generally optimistic and the ministerial direction that, as far as possible, overall results should be similar to previous years.

The extent to which outcomes differ compared to 2019 varies by subject, and some of the differences at subject level are greater than those observed in a typical year. This is due to differences in the percentage of centres with small cohorts, and the approach to awarding grades for centres where the statistical evidence would be too unreliable. Despite this, the grade profiles of individual students have remained broadly stable. This is important when considered in the context of admissions to HE, since this relies on there being general stability within the system from one year to the next.

The following section considers students' results further and focuses on the outcomes for students from different demographic and socio-economic groups.

10 Student-level equalities analysis

10.1 Introduction

To understand the fairness of the approach to awarding grades this summer, it is important to evaluate whether any demographic and socio-economic groups of students have been advantaged or disadvantaged. In Section 7.5 we presented a centre-level equalities analysis which considered the impact of applying the standardisation model using 2019 data. This demonstrated that the standardisation model used did not itself introduce bias into the grading. It did not, however, evaluate the extent of bias in the entire process as it did not consider the rank order information which is a fundamental part of the approach.

This section presents the findings of the student-level analysis undertaken to check the equalities impact of the full approach including the rank orders submitted by centres. This consists of three main strands of work: i) a univariate analysis of each key background variable; ii) an overall multivariate analysis; iii) a specific multivariate analysis of a sample of subjects.

To assess any differential effect of the process on different types of students, we examined the extent to which the relationship between grade outcomes and student background variables in 2018 and 2019 would be maintained in the 2020 outcomes. To ensure like-with-like comparisons, we limited our analyses to data on (i) subjects examined under the same specifications in 2018-2020,[121] (ii) centres with entries in these subjects in each of the years 2018, 2019, 2020, and (iii) students who by 31 August of the respective year reached or will reach the target age of the qualification level of their entries.[122] Table 10.1 shows the number of entries from target-age students, centres and subjects in the resultant datasets.

Table 10.1. Number of entries, centres and subjects in datasets for equalities analysis

	A Level			AS Level		
	Entries	Centres	Subjects	Entries	Centres	Subjects
2018	457464	2547	30	54950	610	44
2019	475296	2547	30	53233	610	44
2020	471229	2547	30	48102	610	44

121 Criterion (i) means that only phase 1 and phase 2 reformed subjects (that is, subjects/specifications that were first assessed in 2017 and 2018 respectively) were included in the analysis. Note that A level mathematics, as a phase 3 reformed subject, was excluded.

122 Centre exclusion was carried out on a subject-by-subject basis. For example, suppose for A Level French, a centre has both 18-year-old and 19-year-old students in each of 2018-2020, and for A Level German, it has both 18-year-old and 19-year-old students in 2018 and 2019 but only 19-year-old students in 2020. Following criterion (iii), data on all three years' 19-year-old students in both languages was excluded, and following criterion (ii), data on all three years' 18-year-old students in French was included and data on the preceding two years' 18-year-old students in German was excluded.

10.2 Student background variables

The datasets were augmented with data on a range of student background variables.

Data on the following variables were taken from the entries data supplied to us by exam boards:

- Gender: each entry was classed as belonging to a male or female student. A very small number of entries had neither male nor female attributed in the data. These entries are not reported in the gender tables due to the very low numbers.
- Prior attainment: a normalised mean GCSE score which can range from 0 to 100 was used as the prior attainment measure. Entries by students with unknown mean GCSE score and entries with out-of-range scores were marked as missing prior attainment data. Entries with non-missing prior attainment data were also classed as belonging to a student with high, medium or low level of prior attainment. To classify students based on their prior attainment, we identified in our dataset all unique students in 2020 with non-missing prior attainment data, and then set the two boundary marks on the normalised mean GCSE score scale that would divide the 2020 students into three groups of roughly equal size of high, medium and low prior attainment. The same two boundary marks were used to class each 2018 and 2019 as well as 2020 entry as belonging to a student with high, medium or low level of prior attainment.

Data on the following background socio-economic and demographic variables were obtained by matching the datasets to extracts of the National Pupil Database (NPD) using students' first name, last name and date of birth as the match key and retaining only the unique matches. Entries by students who could not be uniquely matched or who could be uniquely matched but who had no relevant information in the NPD were marked as missing data on the relevant variable.

- Ethnicity: the EthnicGroupMajor variable in the NPD provided the ethnicity grouping in our analyses. The seven ethnic groups are: AOEG (any other ethnic group), ASIA (Asian), BLAC (Black), CHIN (Chinese), MIXD (mixed background), UNCL (unclassified), WHIT (White).

- Major language: the LanguageGroupMajor variable in the NPD provided the major language grouping used in our analyses. The three major language categories are: ENG (English), OTH (other than English), UNCL (unclassified).

- Special educational needs (SEN): the SENProvisionMajor variable in the NPD provided the SEN provision grouping used in our analyses. The three categories are: NON (no SEN), SNS (SEN without Statement), SS (SEN with Statement). A fourth possible category in the NPD is UNCL (unclassified), but no matched student in our datasets was recorded in the NPD as having unclassified SEN provision.

- Free school meal (FSM): the FSMeligible variable in the NPD provided the FSM eligibility grouping used in our analyses. The two categories are: YES (eligible), NO (not eligible).
- Social economic status (SES): the SES grouping used in our analyses was based on the IDACIScore variable in the NPD. To classify students into SES groups, we identified in our dataset all unique students in 2020 with non-missing IDACI score, and then set the two boundary scores on the IDACI score scale that would divide the 2020 students into three groups of roughly equal size of low, mid and high SES. The same two boundary scores were used to class each 2018 and 2019 as well as 2020 entry as belonging to a student with low, mid or high SES.

10.3 Univariate analysis

To assess the attainment difference between student types, we examined three measures of attainment at the group level, namely, the percentage of entries in the relevant group awarded grade A (in AS level analyses) or grade A and above (in A level analyses), the percentage of entries in the relevant group awarded grade C and above, and the mean of grades awarded for entries in the relevant group.[123]

10.3.1 A level

Tables 10.2 to 10.10 present breakdowns by, respectively, gender, ethnicity, major language, SEN provision status, FSM eligibility status and SES, of the number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in the 2018 and 2019 actual outcomes and the 2020 outcome based on calculated grades. In view of research findings suggesting that high attaining students from disadvantaged backgrounds might be predicted to have lower than actual outcomes[124], further breakdowns by SES are provided separately for students with low, medium and high level of prior attainment in Tables 10.8 to 10.10. To further understand the impact of the process this summer, an additional analysis looking at the differences between CAGs and calculated grades for different socio-economic groups is provided in Annex Q.

123 Letter grades were converted to numbers for the mean grade analysis as follows: A*=6, A=5, …, E=1, U=0.

124 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879605/Equality_impact_assessment_literature_review_15_April_2020.pdf

The descriptive statistics in Table 10.2 show that the mean grades awarded to male and female students this year are very similar to those awarded in 2018 and 2019, as are the percentages of students achieving grade A and grade C. This would be expected as the prior attainment for male and female students is very similar to that of the male and female students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against male or female students.

Table 10.2. A level: Breakdown by students' gender against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
Female	270690	59.2		280013	58.9		279076	59.2	
Male	186771	40.8		195224	41.1		192150	40.8	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Female	94.5	64.81	10.95	94.4	64.87	10.95	95.0	64.49	11.10
Male	91.5	63.04	10.69	91.7	63.01	10.58	92.5	63.45	10.81
Grade A and above	% of group			% of group			% of group		
Female	26.2			25.5			26.5		
Male	24.9			23.4			23.6		
Grade C and above	% of group			% of group			% of group		
Female	79.6			78.9			80.6		
Male	75.8			74.6			74.7		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
Female	3.60	1.35		3.57	1.36		3.63	1.33	
Male	3.48	1.41		3.42	1.41		3.43	1.41	

The descriptive statistics in Table 10.3 show that the mean grades awarded this year to students of different ethnicities are very similar to those awarded in 2018 and 2019, as are the percentages of students of different ethnicities achieving grade A and grade C. This would be expected as the prior attainment for these students is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students from different ethnic backgrounds.

Table 10.3. A level: Breakdown by students' ethnicity against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
AOEG	5502	1.2		6045	1.3		6202	1.3	
ASIA	39704	8.7		45178	9.5		48641	10.3	
BLAC	16763	3.7		18731	3.9		20222	4.3	
CHIN	2369	0.5		2590	0.5		2494	0.5	
MIXD	17022	3.7		18680	3.9		19712	4.2	
UNCL	4817	1.1		5479	1.2		5884	1.2	
WHIT	298212	65.2		300717	63.3		291327	61.8	
Not known	73075	16.0		77876	16.4		76747	16.3	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
AOEG	94.4	63.79	10.89	93.7	63.66	10.78	94.6	63.36	10.60
ASIA	95.0	64.56	10.99	95.6	64.46	10.96	95.9	64.52	11.23
BLAC	94.0	61.63	10.06	94.1	61.66	10.04	94.3	61.72	10.26
CHIN	91.1	69.22	11.43	90.3	69.85	11.82	92.2	69.76	12.09
MIXD	94.7	64.14	10.95	94.6	64.40	11.12	94.9	64.39	11.08
UNCL	93.8	63.64	10.97	93.9	64.23	10.85	93.8	63.26	11.13
WHIT	95.3	64.07	10.76	95.5	64.06	10.70	96.0	64.01	10.85
Not known	83.8	64.48	11.40	83.2	64.55	11.34	84.7	64.54	11.46
Grade A and above	% of group			% of group			% of group		
AOEG	23.4			21.6			22.8		
ASIA	22.8			21.1			22.0		
BLAC	17.3			16.4			17.7		
CHIN	36.6			38.3			39.7		
MIXD	25.7			24.8			24.9		
UNCL	24.1			23.8			23.3		
WHIT	25.7			24.7			25.5		
Not known	28.9			28.1			28.7		
Grade C and above	% of group			% of group			% of group		
AOEG	75.1			73.8			76.6		
ASIA	74.0			72.6			74.0		
BLAC	72.3			70.2			71.2		
CHIN	83.3			84.7			86.0		
MIXD	77.6			77.1			78.2		
UNCL	75.7			76.5			75.3		
WHIT	78.7			77.8			79.1		
Not known	79.3			78.8			79.5		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
AOEG	3.42	1.42		3.36	1.40		3.46	1.36	
ASIA	3.40	1.40		3.33	1.40		3.38	1.38	
BLAC	3.25	1.34		3.18	1.35		3.23	1.35	
CHIN	3.89	1.41		3.96	1.39		4.02	1.34	
MIXD	3.54	1.39		3.52	1.39		3.54	1.36	
UNCL	3.46	1.41		3.48	1.41		3.44	1.40	
WHIT	3.57	1.36		3.53	1.37		3.58	1.35	
Not known	3.65	1.40		3.62	1.40		3.65	1.39	

The descriptive statistics in Table 10.4 show that the mean grades awarded this year to students whose first language was and was not English are very similar to those awarded in 2018 and 2019, as are the percentages of students achieving grade A and grade C. Again, this would be expected as the prior attainment for these students is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students for whom English is the first language.

Table 10.4. A level: Breakdown by students' major language against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
ENG	337911	73.9		344061	72.4		337349	71.6	
OTH	45182	9.9		51647	10.9		55032	11.7	
UNCL	1296	0.3		1712	0.4		2101	0.4	
Not known	73075	16.0		77876	16.4		76747	16.3	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
ENG	95.2	64.12	10.78	95.4	64.15	10.76	95.9	64.07	10.90
OTH	94.5	63.47	10.90	95.0	63.34	10.72	95.3	63.56	10.99
UNCL	92.6	61.77	10.21	93.5	64.42	10.36	92.3	62.52	10.89
Not known	83.8	64.48	11.40	83.2	64.55	11.34	84.7	64.54	11.46
Grade A and above	% of group			% of group			% of group		
ENG	25.5			24.5			25.1		
OTH	21.7			20.3			21.9		
UNCL	18.6			22.7			23.3		
Not known	28.9			28.1			28.7		
Grade C and above	% of group			% of group			% of group		
ENG	78.4			77.5			78.6		
OTH	73.4			72.0			74.3		
UNCL	72.3			77.3			74.3		
Not known	79.3			78.8			79.5		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
ENG	3.56	1.36		3.52	1.37		3.55	1.36	
OTH	3.36	1.40		3.30	1.40		3.39	1.38	
UNCL	3.26	1.42		3.47	1.32		3.42	1.39	
Not known	3.65	1.40		3.62	1.40		3.65	1.39	

The descriptive statistics in Table 10.5 show that the mean grades awarded this year to students with and without special education needs are very similar to those awarded in 2018 and 2019. There is very small drop in the percentage of students with special education needs achieving grade A but not grade C.

Table 10.5. A level: Breakdown by students' SEN provision status against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
NON	365668	79.9		377414	79.4		372853	79.1	
SNS	15028	3.3		16010	3.4		17411	3.7	
SS	3693	0.8		3996	0.8		4218	0.9	
UNCL	0	0.0		0	0.0		0	0.0	
Not known	73075	16.0		77876	16.4		76747	16.3	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
NON	95.2	64.13	10.78	95.4	64.13	10.74	95.9	64.08	10.89
SNS	94.3	62.20	10.73	94.1	62.21	10.91	95.0	62.27	11.03
SS	91.1	62.30	11.40	90.5	62.79	11.09	89.9	62.73	11.29
UNCL	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
Not known	83.8	64.48	11.40	83.2	64.55	11.34	84.7	64.54	11.46
Grade A and above	% of group			% of group			% of group		
NON	25.1			23.9			24.8		
SNS	23.4			23.4			22.0		
SS	26.1			24.8			25.4		
UNCL	0.0			0.0			0.0		
Not known	28.9			28.1			28.7		
Grade C and above	% of group			% of group			% of group		
NON	77.9			76.8			78.1		
SNS	75.6			75.8			75.7		
SS	76.3			76.8			77.1		
UNCL	0.0			0.0			0.0		
Not known	79.3			78.8			79.5		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
NON	3.54	1.37		3.49	1.38		3.54	1.36	
SNS	3.46	1.39		3.46	1.40		3.43	1.38	
SS	3.51	1.42		3.51	1.39		3.53	1.39	
UNCL	0.00	0.00		0.00	0.00		0.00	0.00	
Not known	3.65	1.40		3.62	1.40		3.65	1.39	

The descriptive statistics in Table 10.6 show that the mean grades awarded this year to students who are and are not eligible for free school meals are very similar to those awarded in 2018 and 2019, as are the percentages of students achieving grade A and grade C. This would be expected as the prior attainment for students who are and are not eligible for free school meals is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students who are and are not eligible for free school meals.

Table 10.6. A level: Breakdown by students' FSM eligibility status against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
NO	367330	80.3		376370	79.2		370649	78.7	
YES	17059	3.7		21050	4.4		23833	5.1	
Not known	73075	16.0		77876	16.4		76747	16.3	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
NO	95.1	64.16	10.79	95.3	64.18	10.76	95.8	64.17	10.92
YES	94.1	61.37	10.49	94.8	61.65	10.34	95.4	61.13	10.35
Not known	83.8	64.48	11.40	83.2	64.55	11.34	84.7	64.54	11.46
Grade A or above	% of group			% of group			% of group		
NO	25.3			24.2			25.1		
YES	18.6			18.1			18.0		
Not known	28.9			28.1			28.7		
Grade C and above	% of group			% of group			% of group		
NO	78.1			77.1			78.3		
YES	71.2			70.9			71.9		
Not known	79.3			78.8			79.5		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
NO	3.55	1.37		3.50	1.38		3.55	1.36	
YES	3.25	1.39		3.24	1.37		3.26	1.35	
Not known	3.65	1.40		3.62	1.40		3.65	1.39	

The descriptive statistics in Table 10.7 show that the mean grades awarded this year to students of different socio-economic status are very similar to those awarded in 2018 and 2019, as are the percentages of students of different socio-economic status achieving grade A and grade C. This would be expected as the prior attainment for these students is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students from different socio-economic backgrounds.

Table 10.7. A level: Breakdown by student's SES against number against entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
Low	120979	26.4		127219	26.8		128826	27.3	
Mid	130427	28.5		134985	28.4		131277	27.9	
High	132181	28.9		134479	28.3		133647	28.4	
Not known	73877	16.1		78613	16.5		77479	16.4	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	95.0	62.66	10.59	95.1	62.75	10.50	95.6	62.68	10.65
Mid	95.2	64.05	10.72	95.3	64.10	10.73	96.0	63.99	10.86
High	95.1	65.29	10.90	95.6	65.21	10.87	95.8	65.25	11.06
Not known	83.9	64.45	11.40	83.3	64.53	11.33	84.8	64.53	11.45
Grade A and above	% of group			% of group			% of group		
Low	21.0			20.0			21.0		
Mid	25.1			24.0			24.6		
High	28.7			27.6			28.3		
Not known	28.9			28.0			28.7		
Grade C and above	% of group			% of group			% of group		
Low	74.0			72.6			74.6		
Mid	78.0			77.2			78.2		
High	81.1			80.3			81.0		
Not known	79.3			78.8			79.4		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
Low	3.37	1.38		3.31	1.38		3.38	1.36	
Mid	3.54	1.37		3.50	1.37		3.54	1.36	
High	3.68	1.35		3.64	1.36		3.67	1.35	
Not known	3.65	1.40		3.62	1.40		3.65	1.39	

Tables 10.8, 10.9 and 10.10 show the descriptive statistics for students from different socio-economic backgrounds split by prior attainment (low, medium and high).

Split by prior attainment, the mean grades awarded this year to students from different socio-economic backgrounds are very similar to those awarded in 2018 and 2019, as are the percentages of students of different backgrounds achieving grade A and grade C. The observed small fluctuations in outcomes are akin to the fluctuations between 2018 and 2019 and as such do not suggest that the process of awarding grades has been biased for or against students from different socio-economic backgrounds. There is a small increase in the proportion of students achieving grade A but this is regardless of socio-economic background (Table 10.10).

Table 10.8. A level: Breakdown by SES of students with low prior attainment against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number		% of all	Number		% of all	Number		% of all
Low	38233		30.7	40634		31.2	42159		31.7
Mid	35745		28.7	37403		28.7	37882		28.5
High	32165		25.8	33151		25.4	33629		25.3
Not known	18531		14.9	19187		14.7	19440		14.6
Prior attainment[125]	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	100.0	51.50	4.78	100.0	51.75	4.45	100.0	51.79	4.51
Mid	100.0	51.87	4.52	100.0	51.98	4.26	100.0	52.12	4.29
High	100.0	52.18	4.26	100.0	52.27	4.11	100.0	52.37	4.14
Not known	100.0	51.80	4.45	100.0	51.95	4.21	100.0	51.98	4.21
Grade A and above	% of group			% of group			% of group		
Low	4.3			3.7			3.7		
Mid	4.3			3.6			3.6		
High	4.9			4.2			3.7		
Not known	5.7			5.3			4.6		
Grade C and above	% of group			% of group			% of group		
Low	53.7			51.8			54.7		
Mid	55.4			53.9			56.2		
High	58.1			56.3			56.9		
Not known	57.4			56.5			56.8		
Mean grade	Mean		SD	Mean		SD	Mean		SD
Low	2.57		1.20	2.51		1.20	2.58		1.16
Mid	2.62		1.18	2.56		1.18	2.61		1.13
High	2.69		1.17	2.63		1.18	2.64		1.12
Not known	2.69		1.22	2.66		1.22	2.66		1.17

125 To clarify, note that references to ‘low’, ‘medium’ and ‘high’ under these headings relate to the levels of SES. All students in Table 10.8 are in the ‘low’ prior-attainment category, with ‘medium’ and ‘high’ prior-attainment students included in Tables 10.9 and 10.10 respectively. This can be seen from differences in mean prior-attainment across these 3 tables.

Table 10.9. A level: Breakdown by SES of students with medium prior attainment against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
Low	40066	27.8		42426	27.8		43021	28.3	
Mid	42781	29.7		44591	29.2		43797	28.8	
High	41501	28.8		43929	28.7		43539	28.6	
Not known	19611	13.6		21906	14.3		21799	14.3	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	100.0	62.17	2.73	100.0	62.28	2.74	100.0	62.27	2.71
Mid	100.0	62.21	2.74	100.0	62.32	2.73	100.0	62.31	2.71
High	100.0	62.28	2.73	100.0	62.37	2.72	100.0	62.42	2.69
Not known	100.0	62.17	2.71	100.0	62.31	2.72	100.0	62.33	2.71
Grade A and above	% of group			% of group			% of group		
Low	12.9			11.8			12.2		
Mid	14.5			13.2			13.0		
High	15.0			14.8			13.6		
Not known	15.4			14.8			14.5		
Grade C and above	% of group			% of group			% of group		
Low	75.8			73.8			76.4		
Mid	78.5			77.5			79.2		
High	80.3			79.8			81.2		
Not known	77.8			77.8			79.1		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
Low	3.26	1.19		3.19	1.20		3.27	1.17	
Mid	3.35	1.17		3.31	1.16		3.35	1.14	
High	3.41	1.14		3.39	1.15		3.40	1.11	
Not known	3.36	1.20		3.35	1.19		3.37	1.16	

Table 10.10. A level: Breakdown by SES of students with high prior attainment against number of entries, students' prior attainment, percentage of grade A and above, percentage of grade C and above and mean grade in 2018-2020 A level outcomes

	2018			2019			2020		
Entries	Number		% of all	Number		% of all	Number		% of all
Low	36582		23.1	37904		23.6	37915		24.1
Mid	45598		28.8	46626		29.1	44398		28.2
High	52054		32.9	51417		32.1	50811		32.2
Not known	23842		15.1	24405		15.2	24435		15.5
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	100.0	74.88	6.38	100.0	75.05	6.39	100.0	74.88	6.38
Mid	100.0	75.33	6.56	100.0	75.54	6.67	100.0	75.33	6.56
High	100.0	75.79	6.78	100.0	75.98	6.88	100.0	75.79	6.78
Not known	100.0	76.17	7.09	100.0	76.40	7.38	100.0	76.17	7.09
Grade A and above	% of group			% of group			% of group		
Low	45.3			44.8			48.6		
Mid	49.2			48.8			52.6		
High	52.7			52.1			55.8		
Not known	51.9			51.1			54.6		
Grade C and above	% of group			% of group			% of group		
Low	92.5			92.6			93.9		
Mid	94.1			94.6			95.2		
High	95.3			95.5			96.1		
Not known	94.4			94.1			95.1		
Mean grade	Mean		SD	Mean		SD	Mean		SD
Low	4.26		1.17	4.25		1.15	4.36		1.13
Mid	4.37		1.13	4.38		1.11	4.47		1.09
High	4.47		1.09	4.47		1.08	4.55		1.06
Not known	4.44		1.13	4.42		1.14	4.51		1.12

Attainment gaps on the various student background variables can be calculated from the tables above. By examining the extent to which attainment gaps in the 2020 grade outcomes have changed relative to the attainment gaps seen in previous years' grade outcomes, we can assess any differential effect of the 2020 process of grading and awarding on different groups of students.

Table 10.11 gives a summary of the attainment gaps on each attainment measure that can be calculated from the numbers presented in Tables 10.2 to 10.7. For all but the SES variable, the attainment gap was calculated by subtracting the outcome of the group with fewer entries from the outcome of the group with more entries. Hence, a positive number indicates higher performance of the majority group while a negative number indicates lower performance of the majority group. For the SES variable, the attainment gap was calculated by subtracting the outcome of the low SES group from that of the high SES group and therefore a positive number indicates higher outcomes for the high SES group. When two or more groups were combined to be contrasted with another group, a weighted average was calculated for the composite group. For example, under FSM, a weighted average of the NO and YES groups was calculated to be compared with the 'unknown' group; under Ethnicity, a weighted average of the AOEG, ASIA, BLAC, CHIN, MIXD and WHIT groups was calculated to be compared with the weighted average of the UNCL and 'unknown' groups.

To illustrate how to read Table 10.11, the first row is considered as an example. The first row shows: in 2018 the proportion of entries by female students receiving grade A was higher than the proportion of entries by male students receiving grade A. This gap was 1.3 percentage points. The attainment gap continued in 2019, widening by 0.8 percentage point to 2.1 percentage points. In 2020 the attainment gap is 2.8 percentage points, which represents an increase of 0.7 percentage points from 2019 and 1.1 percentage points from the weighted average attainment gap across 2018 and 2019.

As the '19-18 Difference' column in Table 10.11 illustrates, attainment gaps seldom stay constant from year-to-year. The size of the changes in attainment gaps observed between 2019 and 2020 are similar to those seen between 2018 and 2019. For gender, there are fluctuations that are similar to those seen over time with a slight improvement in outcomes for females compared to males, and as will be shown in the multivariate analyses, this is marginal in terms of effect size in the context of the cautious thresholds applied below. The changes do not suggest that the process of awarding grades this summer has been biased, rather the changes are akin to normal year on year fluctuations.

Table 10.11. A level: Attainment gaps in 2018 and 2019 outcomes, differences between 2018 and 2019 attainment gaps, attainment gaps in 2020 outcome and differences between 2020 attainment gaps from weighted average attainment gaps of 2018 and 2019

	2018	2019	19-18	2020	
	Outcome	**Outcome**	**Difference**	**Outcome**	**Difference**
Percentage of grade A and above (Difference shown is percentage point difference)					
Gender					
Female – Male	1.3	2.1	0.8	2.8	1.1
Ethnicity					
WHIT – AOEG	2.3	3.0	0.7	2.7	0.0
WHIT – ASIA	2.9	3.5	0.6	3.5	0.3
WHIT – BLAC	8.4	8.3	-0.1	7.8	-0.5
WHIT – CHIN	-11.0	-13.6	-2.6	-14.2	-1.9
WHIT – MIXD	0.0	-0.2	-0.1	0.6	0.7
Known – (unknown+UNCL)	-3.6	-3.9	-0.4	-3.7	0.1
Major language					
ENG – OTH	3.8	4.2	0.4	3.2	-0.8
Known – (unknown+UNCL)	-3.7	-4.1	-0.4	-3.9	0.0
SEN					
NON – (SNS+SS)	1.2	0.2	-1.0	2.1	1.4
Known – (unknown+UNCL)	-3.9	-4.2	-0.3	-4.1	0.0
FSM					
NON – YES	6.8	6.1	-0.6	7.1	0.7
Known – unknown	-3.9	-4.2	-0.3	-4.1	0.0
SES					
High – Low	7.8	7.6	-0.2	7.3	-0.4
Known – unknown	-3.8	-4.1	-0.3	-4.0	0.0

	2018	2019	19-18	2020	
	Outcome	**Outcome**	**Difference**	**Outcome**	**Difference**
Percentage of grade C and above (Difference shown is percentage point difference)					
Gender					
Female – Male	3.8	4.3	0.5	5.9	1.9
Ethnicity					
WHIT – AOEG	3.6	4.0	0.4	2.5	-1.3
WHIT – ASIA	4.7	5.2	0.5	5.0	0.1
WHIT – BLAC	6.4	7.6	1.2	7.8	0.8
WHIT – CHIN	-4.6	-6.9	-2.3	-6.9	-1.2
WHIT – MIXD	1.1	0.7	-0.4	0.9	0.0
Known – (unknown+UNCL)	1.3	1.9	0.6	1.2	-0.4
Major language					
ENG – OTH	5.0	5.5	0.4	4.2	-1.0
Known – (unknown+UNCL)	1.4	2.0	0.6	1.4	-0.3
SEN					
NON – (SNS+SS)	2.2	0.8	-1.3	2.1	0.6
Known – (unknown+UNCL)	1.5	2.0	0.5	1.5	-0.3
FSM					
NON – YES	6.9	6.3	-0.7	6.4	-0.2
Known – unknown	1.5	2.0	0.5	1.5	-0.3
SES					
High – Low	7.1	7.6	0.6	6.4	-1.0
Known – unknown	1.5	2.0	0.5	1.5	-0.3

	2018	2019	19-18	2020	
	Outcome	Outcome	Difference	Outcome	Difference
Mean grade on 0-6 scale (Difference shown is grade point difference)					
Gender					
Female – Male	0.12	0.15	0.03	0.20	0.06
Ethnicity					
WHIT – AOEG	0.15	0.16	0.02	0.12	-0.04
WHIT – ASIA	0.17	0.20	0.03	0.19	0.01
WHIT – BLAC	0.32	0.34	0.02	0.34	0.01
WHIT – CHIN	-0.32	-0.43	-0.11	-0.45	-0.07
WHIT – MIXD	0.03	0.01	-0.01	0.04	0.02
Known – (unknown+UNCL)	0.10	0.12	0.02	0.11	-0.01
Major language					
ENG – OTH	0.19	0.22	0.02	0.16	-0.04
Known – (unknown+UNCL)	0.11	0.13	0.02	0.12	0.00
SEN					
NON – (SNS+SS)	0.07	0.02	-0.05	0.09	0.04
Known – (unknown+UNCL)	0.11	0.13	0.02	0.12	0.00
FSM					
NON – YES	0.30	0.27	-0.03	0.29	0.01
Known – unknown	0.11	0.13	0.02	0.12	0.00
SES					
High – Low	0.32	0.33	0.01	0.29	-0.04
Known – unknown	0.11	0.13	0.02	0.12	0.00

10.3.2 AS level

Tables 10.12 to 10.20 present breakdowns by, respectively, student's gender, ethnicity, major language, SEN provision status, FSM eligibility status and SES, of the number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in the 2018 and 2019 actual outcomes and the 2020 outcome based on calculated grades. Tables 10.18 to 10.20 provide further breakdowns by SES for students with low, medium and high level of prior attainment separately.

It should be noted that the comparisons drawn at AS are less reliable than those made for other qualifications. This is due to the significant reductions in entry that have occurred over recent years. While care has been taken to find a common profile of centres between years, changes in entry strategy have also taken place within centres to which this centre-level matching will not be sensitive. Fluctuations in attainment gaps are therefore more likely to occur regardless of the process of awarding grades.

The descriptive statistics in Table 10.12 show that the mean grades awarded to male and female students this year are very similar to those awarded in 2018 and 2019, as are the percentages of students achieving grade A and grade C. This would be expected as the prior attainment for male and female students is very similar to that of the male and female students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against male or female students.

Table 10.12. AS level: Breakdown by student's gender against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
Female	29963	54.5		28652	53.8		25939	53.9	
Male	24987	45.5		24581	46.2		22161	46.1	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Female	93.6	63.64	11.07	94.2	63.75	11.02	94.8	63.50	11.25
Male	92.1	63.44	11.06	92.1	64.30	11.68	93.5	63.97	11.51
Grade A	% of group			% of group			% of group		
Female	19.0			20.0			20.1		
Male	23.1			23.8			22.4		
Grade C and above	% of group			% of group			% of group		
Female	58.9			61.3			62.6		
Male	58.7			59.7			57.9		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
Female	2.81	1.61		2.89	1.59		2.94	1.56	
Male	2.86	1.68		2.88	1.69		2.82	1.68	

The descriptive statistics in Table 10.13 show that the mean grades awarded this year to students of different ethnicities are very similar to those awarded in 2018 and 2019, as are the percentages of students of different ethnicities achieving grade A. This would be expected as the prior attainment for these students is very similar to that of the students entered in 2018 and 2019. There are some larger fluctuations at grade C, however where these occur there is evidence that they also occurred between 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students from different ethnic backgrounds.

Table 10.13. AS level: Breakdown by students' ethnicity against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number		% of all	Number		% of all	Number		% of all
AOEG	1067		1.9	982		1.8	867		1.8
ASIA	7863		14.3	8173		15.4	7510		15.6
BLAC	3527		6.4	3594		6.8	3257		6.8
CHIN	383		0.7	365		0.7	325		0.7
MIXD	2381		4.3	2441		4.6	2210		4.6
UNCL	698		1.3	574		1.1	602		1.3
WHIT	29776		54.2	27988		52.6	24641		51.2
Not known	9255		16.8	9116		17.1	8690		18.1
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
AOEG	94.7	64.77	10.95	94.9	64.79	10.95	97.2	64.66	11.34
ASIA	95.0	64.91	11.47	95.3	65.85	12.17	96.1	65.91	12.00
BLAC	94.4	62.52	10.36	94.5	62.35	10.76	96.0	62.39	10.98
CHIN	94.3	72.79	12.41	93.2	73.07	11.73	94.5	71.13	12.34
MIXD	93.4	63.71	11.47	94.6	65.02	12.03	95.8	64.10	11.72
UNCL	90.0	62.78	11.07	88.3	63.21	11.08	93.4	62.77	11.23
WHIT	96.2	63.66	11.06	96.3	63.90	11.15	96.8	63.63	11.22
Not known	80.1	61.63	10.34	81.2	62.38	10.57	84.0	61.92	10.76
Grade A	% of group			% of group			% of group		
AOEG	21.6			23.6			22.8		
ASIA	23.0			25.0			24.4		
BLAC	17.2			16.6			17.9		
CHIN	47.3			43.6			42.8		
MIXD	20.3			24.1			22.9		
UNCL	17.9			20.6			17.8		
WHIT	20.7			21.1			20.7		
Not known	20.2			21.2			19.8		
Grade C and above	% of group			% of group			% of group		
AOEG	61.0			62.8			65.2		
ASIA	59.4			62.0			62.8		
BLAC	55.7			55.3			56.7		
CHIN	77.8			81.4			75.1		
MIXD	58.1			63.0			62.5		
UNCL	56.9			62.0			58.8		
WHIT	59.4			60.5			61.1		
Not known	56.9			59.4			56.6		
Mean grade	Mean		SD	Mean		SD	Mean		SD
AOEG	2.91		1.61	2.95		1.64	3.04		1.57
ASIA	2.88		1.67	2.96		1.68	2.99		1.63
BLAC	2.69		1.62	2.68		1.62	2.73		1.63
CHIN	3.70		1.57	3.78		1.38	3.62		1.55
MIXD	2.80		1.64	2.98		1.63	2.97		1.61
UNCL	2.77		1.59	2.86		1.64	2.81		1.59
WHIT	2.85		1.63	2.89		1.62	2.90		1.60
Not known	2.76		1.66	2.83		1.65	2.75		1.65

The descriptive statistics in Table 10.14 show that the mean grades awarded this year to students whose first language was and was not English are very similar to those awarded in 2018 and 2019, as are the percentages of students achieving grade A and grade C. Where fluctuations occur they also occurred between 2018 and2019. Again, this broad stability would be expected as the prior attainment for these students is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students for whom English is the first language.

Table 10.14. AS level: Breakdown by students' major language against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
ENG	36006	65.5		34161	64.2		30276	62.9	
OTH	9469	17.2		9749	18.3		8935	18.6	
UNCL	220	0.4		207	0.4		201	0.4	
Not known	9255	16.8		9116	17.1		8690	18.1	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
ENG	96.0	63.83	11.22	96.4	64.18	11.38	97.0	63.88	11.41
OTH	94.1	63.99	10.92	93.6	64.69	11.63	94.6	64.76	11.61
UNCL	88.6	67.07	9.40	88.4	62.40	9.99	90.1	61.41	9.42
Not known	80.1	61.63	10.34	81.2	62.38	10.57	84.0	61.92	10.76
Grade A	% of group			% of group			% of group		
ENG	21.1			21.7			21.1		
OTH	20.3			22.6			22.9		
UNCL	24.5			15.9			12.9		
Not known	20.2			21.2			19.8		
Grade C and above	% of group			% of group			% of group		
ENG	59.3			61.1			61.2		
OTH	58.6			59.9			61.8		
UNCL	67.3			51.2			56.2		
Not known	56.9			59.4			56.6		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
ENG	2.85	1.63		2.91	1.62		2.91	1.61	
OTH	2.82	1.64		2.87	1.67		2.94	1.62	
UNCL	3.12	1.50		2.49	1.69		2.70	1.47	
Not known	2.76	1.66		2.83	1.65		2.75	1.65	

The descriptive statistics in Table 10.15 show that the mean grades awarded this year to students with and without special education needs are very similar to those awarded in 2018 and 2019, as are the percentages of students with and without special educational needs achieving grade A and grade C. Where small fluctuations occur they also occurred between 2018 and 2019. This broad stability would be expected as the prior attainment for these students is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students with special educational needs.

Table 10.15. AS level: Breakdown by students' SEN provision status against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
NON	43384	79.0		41950	78.8		36805	76.5	
SNS	1853	3.4		1757	3.3		1937	4.0	
SS	458	0.8		410	0.8		670	1.4	
UNCL	0	0.0		0	0.0		0	0.0	
Not known	9255	16.8		9116	17.1		8690	18.1	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
NON	95.5	64.05	11.13	95.8	64.44	11.42	96.5	64.27	11.45
SNS	95.7	60.71	11.32	96.0	61.24	11.34	96.1	60.89	11.17
SS	96.1	60.43	10.49	92.9	61.51	10.62	96.6	61.97	11.12
UNCL	0.0	0.00	0.00	0.0	0.00	0.00	0.0	0.00	0.00
Not known	80.1	61.63	10.34	81.2	62.38	10.57	84.0	61.92	10.76
Grade A	% of group			% of group			% of group		
NON	21.1			22.1			21.8		
SNS	17.6			17.5			15.5		
SS	19.7			17.1			21.8		
UNCL	0.0			0.0			0.0		
Not known	20.2			21.2			19.8		
Grade C and above	% of group			% of group			% of group		
NON	59.5			60.9			61.7		
SNS	52.2			57.3			54.2		
SS	56.8			59.0			60.0		
UNCL	0.0			0.0			0.0		
Not known	56.9			59.4			56.6		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
NON	2.85	1.63		2.91	1.63		2.93	1.61	
SNS	2.64	1.63		2.72	1.63		2.62	1.59	
SS	2.82	1.60		2.78	1.62		2.88	1.63	
UNCL	0.00	0.00		0.00	0.00		0.00	0.00	
Not known	2.76	1.66		2.83	1.65		2.75	1.65	

The descriptive statistics in Table 10.16 show that the mean grades awarded this year to students who are and are not eligible for free school meals are very similar to those awarded in 2018 and 2019, as are the percentages of students achieving grade A and grade C. Where fluctuations occur they also occurred between 2018 and 2019. This stability would be expected as the prior attainment for students who are and are not eligible for free school meals is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students who are and are not eligible for free school meals.

Table 10.16. AS level: Breakdown by student's FSM eligibility status against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
NO	43045	78.3		40918	76.9		36288	75.4	
YES	2650	4.8		3199	6.0		3124	6.5	
Not known	9255	16.8		9116	17.1		8690	18.1	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
NO	95.6	64.09	11.16	95.7	64.55	11.47	96.4	64.42	11.45
YES	94.6	60.39	10.47	96.2	60.94	10.39	97.1	59.94	10.64
Not known	80.1	61.63	10.34	81.2	62.38	10.57	84.0	61.92	10.76
Grade A	% of group			% of group			% of group		
NO	21.4			22.4			22.1		
YES	13.4			15.3			14.6		
Not known	20.2			21.2			19.8		
Grade C and above	% of group			% of group			% of group		
NO	59.8			61.4			61.9		
YES	49.2			52.7			54.8		
Not known	56.9			59.4			56.6		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
NO	2.87	1.63		2.92	1.63		2.94	1.61	
YES	2.46	1.61		2.57	1.63		2.60	1.58	
Not known	2.76	1.66		2.83	1.65		2.75	1.65	

The descriptive statistics in Table 10.17 show that the mean grades awarded this year to students of different socio-economic status are very similar to those awarded in 2018 and 2019, as are the percentages of students of different socio-economic status achieving grade A and grade C. This would be expected as the prior attainment for these students is very similar to that of the students entered in 2018 and 2019. There is no evidence that the process of awarding grades has been biased for or against students from different socio-economic backgrounds.

Table 10.17. AS level: Breakdown by student's SES against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number		% of all	Number		% of all	Number		% of all
Low	15512		28.2	15010		28.2	13704		28.5
Mid	14768		26.9	14412		27.1	13169		27.4
High	15340		27.9	14625		27.5	12455		25.9
Not known	9330		17.0	9186		17.3	8774		18.2
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	95.0	62.25	10.62	95.2	62.94	10.94	96.4	62.63	11.3
Mid	96.1	64.14	11.2	96.2	64.29	11.56	96.5	64.04	11.23
High	95.7	65.27	11.41	95.9	65.64	11.63	96.5	65.68	11.65
Not known	80.2	61.62	10.37	81.2	62.41	10.57	84.1	61.91	10.74
Grade A	% of group			% of group			% of group		
Low	16.9			18.8			18.4		
Mid	21.9			22.0			20.9		
High	24.2			24.9			25.5		
Not known	20.2			21.2			19.7		
Grade C and above	% of group			% of group			% of group		
Low	54.7			57.3			57.8		
Mid	60.2			61.0			61.2		
High	62.8			64.1			65.3		
Not known	56.9			59.5			56.7		
Mean grade	Mean		SD	Mean		SD	Mean		SD
Low	2.66		1.62	2.75		1.64	2.77		1.61
Mid	2.88		1.64	2.90		1.64	2.92		1.59
High	3.00		1.62	3.05		1.61	3.07		1.61
Not known	2.76		1.66	2.84		1.65	2.75		1.65

Tables 10.18, 10.19 and 10.20 show the descriptive statistics for students from different socio-economic backgrounds split by prior attainment (low, medium and high). Split by prior attainment, the mean grades awarded this year to students from different socio-economic backgrounds are very similar to those awarded in 2018 and 2019, as are the percentages of students of different backgrounds achieving grade A and grade C. The observed small fluctuations in outcomes are akin to the fluctuations between 2018 and 2019 and as such do not suggest that the process of awarding grades has been biased for or against students from different socio-economic backgrounds.

Table 10.18. AS level: Breakdown by SES of students with low prior attainment against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number		% of all	Number		% of all	Number		% of all
Low	5128		32.3	4555		31.2	4520		32.8
Mid	4174		26.3	4022		27.5	3636		26.4
High	3807		24.0	3461		23.7	2971		21.5
Not known	2746		17.3	2567		17.6	2673		19.4
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	100.0	51.17	4.60	100.0	51.09	4.61	100.0	50.90	4.42
Mid	100.0	51.55	4.26	100.0	51.31	4.28	100.0	51.29	4.43
High	100.0	51.52	4.20	100.0	51.59	4.32	100.0	51.49	4.28
Not known	100.0	51.51	4.17	100.0	51.82	3.80	100.0	51.44	4.06
Grade A	% of group			% of group			% of group		
Low	2.5			2.5			2.1		
Mid	2.7			2.4			2.2		
High	2.5			3.1			2.2		
Not known	3.4			2.4			2.8		
Grade C and above	% of group			% of group			% of group		
Low	29.1			30.1			30.2		
Mid	30.9			32.0			31.2		
High	31.8			30.9			32.4		
Not known	33.7			32.4			30.9		
Mean grade	Mean		SD	Mean		SD	Mean		SD
Low	1.70		1.37	1.71		1.39	1.73		1.34
Mid	1.74		1.38	1.75		1.39	1.80		1.33
High	1.79		1.38	1.79		1.39	1.80		1.35
Not known	1.86		1.42	1.76		1.39	1.78		1.36

Table 10.19. AS level: Breakdown by SES of students with medium prior attainment against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number	% of all		Number	% of all		Number	% of all	
Low	5318	29.0		5341	29.6		4632	28.6	
Mid	5005	27.3		5007	27.7		4664	28.8	
High	5115	27.9		4909	27.2		4180	25.8	
Not known	2876	15.7		2800	15.5		2703	16.7	
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	100.0	62.45	2.97	100.0	62.44	3.07	100.0	62.21	3.08
Mid	100.0	62.34	3.04	100.0	62.44	3.06	100.0	62.33	3.07
High	100.0	62.43	2.97	100.0	62.35	3.05	100.0	62.39	3.06
Not known	100.0	62.15	3.00	100.0	62.04	3.01	100.0	61.90	3.05
Grade A	%			%			%		
Low	10.7			10.7			9.7		
Mid	11.6			11.5			9.7		
High	11.6			10.1			10.7		
Not known	10.3			11.4			9.1		
Grade C and above	%			%			%		
Low	55.2			55.4			57.2		
Mid	57.0			57.2			58.1		
High	57.0			58.0			58.9		
Not known	54.9			57.0			54.4		
Mean grade	Mean	SD		Mean	SD		Mean	SD	
Low	2.63	1.45		2.62	1.48		2.67	1.41	
Mid	2.68	1.47		2.68	1.47		2.70	1.40	
High	2.70	1.45		2.70	1.40		2.70	1.44	
Not known	2.58	1.50		2.67	1.46		2.55	1.48	

Table 10.20. AS level: Breakdown by SES of students with high prior attainment against number of entries, students' prior attainment, percentage of grade A, percentage of grade C and above and mean grade in 2018-2020 AS level outcomes

	2018			2019			2020		
Entries	Number		% of all	Number		% of all	Number		% of all
Low	4282		25.3	4389		25.9	4053		26.4
Mid	5009		29.6	4839		28.5	4407		28.8
High	5756		34.1	5658		33.3	4869		31.8
Not known	1857		11.0	2089		12.3	2002		13.1
Prior attainment	%known	Mean	SD	%known	Mean	SD	%known	Mean	SD
Low	100.0	75.27	5.91	100.0	75.83	6.63	100.0	76.20	6.77
Mid	100.0	76.43	6.68	100.0	77.00	7.28	100.0	76.36	6.95
High	100.0	76.89	6.70	100.0	77.08	7.40	100.0	77.16	7.22
Not known	100.0	75.75	6.47	100.0	75.92	6.94	100.0	75.90	6.93
Grade A	%			%			%		
Low	41.6			44.5			46.0		
Mid	47.2			48.5			47.2		
High	48.9			50.3			52.2		
Not known	42.0			44.2			45.7		
Grade C and above	%			%			%		
Low	84.7			87.3			88.4		
Mid	87.7			88.6			88.8		
High	87.7			88.9			90.7		
Not known	82.8			86.3			86.2		
Mean grade	Mean		SD	Mean		SD	Mean		SD
Low	3.85		1.27	3.96		1.21	4.00		1.20
Mid	4.01		1.19	4.05		1.17	4.05		1.16
High	4.03		1.21	4.09		1.15	4.16		1.12
Not known	3.81		1.34	3.95		1.21	3.96		1.25

Table 10.21 gives a summary of the attainment gaps on each attainment measure that can be calculated from the numbers presented in Tables 10.12 to 10.20. As the '19-18 Difference' column in Table 10.21 again illustrates, most attainment gaps vary between years. The size of the changes in attainment gaps observed between 2019 and 2020 are similar to those seen between 2018 and 2019. As with A level, for gender, there is some improvement for females relative to males and as will be shown below in the multivariate analyses, this is marginal in terms of effect size compared to the cautious threshold used. The changes do not suggest that the process of awarding grades this summer has been biased, rather the changes are similar to normal year on year fluctuations.

Table 10.21.AS level: Attainment gaps in 2018 and 2019 outcomes, differences between 2018 and 2019 attainment gaps, attainment gaps in 2020 outcome and differences between 2020 attainment gaps from weighted average attainment gaps of 2018 and 2019

	2018	2019	19-18	2020	
	Outcome	**Outcome**	**Difference**	**Outcome**	**Difference**
Percentage of grade A (Difference shown is percentage point difference)					
Gender					
Female – Male	-4.2	-3.7	0.5	-2.3	1.7
Ethnicity					
WHIT – AOEG	-1.0	-2.5	-1.5	-2.1	-0.4
WHIT – ASIA	-2.3	-3.9	-1.6	-3.7	-0.6
WHIT – BLAC	3.5	4.5	1.0	2.8	-1.2
WHIT – CHIN	-26.6	-22.4	4.2	-22.1	2.5
WHIT – MIXD	0.4	-3.0	-3.4	-2.2	-0.9
Known – (unknown+UNCL)	1.0	0.8	-0.2	1.9	1.0
Major language					
ENG – OTH	0.8	-0.8	-1.7	-1.7	-1.8
Known – (unknown+UNCL)	0.6	0.9	0.2	1.9	1.2
SEN					
NON – (SNS+SS)	3.1	4.7	1.6	4.7	0.8
Known – (unknown+UNCL)	0.7	0.7	0.0	1.7	1.0
FSM					
NON – YES	8.1	7.1	-0.9	7.5	-0.1
Known – unknown	0.7	0.7	0.0	1.7	1.0
SES					
High – Low	7.3	6.0	-1.3	7.1	0.5
Known – unknown	0.7	0.7	-0.1	1.8	1.1

	2018	2019	19-18	2020	
	Outcome	**Outcome**	**Difference**	**Outcome**	**Difference**
Percentage of grade C and above (Difference shown is percentage point difference)					
Gender					
Female – Male	0.2	1.6	1.4	4.7	3.8
Ethnicity					
WHIT – AOEG	-1.6	-2.3	-0.7	-4.0	-2.1
WHIT – ASIA	0.0	-1.5	-1.5	-1.6	-0.9
WHIT – BLAC	3.7	5.3	1.6	4.5	0.0
WHIT – CHIN	-18.4	-20.8	-2.4	-13.9	5.6
WHIT – MIXD	1.3	-2.5	-3.7	-1.4	-0.9
Known – (unknown+UNCL)	-2.3	-1.2	1.1	-4.6	-2.8
Major language					
ENG – OTH	0.7	1.1	0.4	-0.5	-1.5
Known – (unknown+UNCL)	-2.0	-1.6	0.4	-4.8	-2.9
SEN					
NON – (SNS+SS)	6.4	3.3	-3.1	6.1	1.2
Known – (unknown+UNCL)	-2.3	-1.4	0.9	-4.7	-2.9
FSM					
NON – YES	10.6	8.7	-1.9	7.1	-2.6
Known – unknown	-2.3	-1.4	0.9	-4.7	-2.9
SES					
High – Low	8.1	6.8	-1.4	7.5	0.0
Known – unknown	-2.3	-1.3	1.0	-4.6	-2.8

	2018	2019	19-18	2020	
	Outcome	**Outcome**	**Difference**	**Outcome**	**Difference**
Mean grade on 0-5 scale (Difference shown is grade point difference)					
Gender					
Female – Male	-0.04	0.02	0.06	0.12	0.13
Ethnicity					
WHIT – AOEG	-0.07	-0.06	0.00	-0.14	-0.08
WHIT – ASIA	-0.03	-0.08	-0.05	-0.09	-0.04
WHIT – BLAC	0.16	0.20	0.04	0.17	-0.01
WHIT – CHIN	-0.85	-0.90	-0.04	-0.72	0.16
WHIT – MIXD	0.05	-0.10	-0.14	-0.07	-0.05
Known – (unknown+UNCL)	-0.09	-0.06	0.02	-0.16	-0.09
Major language					
ENG – OTH	0.03	0.04	0.01	-0.04	-0.07
Known – (unknown+UNCL)	-0.07	-0.07	0.00	-0.16	-0.09
SEN					
NON – (SNS+SS)	0.18	0.18	0.00	0.24	0.07
Known – (unknown+UNCL)	-0.08	-0.06	0.02	-0.16	-0.09
FSM					
NON – YES	0.41	0.35	-0.06	0.35	-0.04
Known – unknown	-0.08	-0.06	0.02	-0.16	-0.09
SES					
High – Low	0.34	0.30	-0.04	0.30	-0.02
Known – unknown	-0.08	-0.06	0.02	-0.16	-0.09

10.4 Multivariate analysis: across subjects

Tables 10.2 to 10.21 provide univariate analyses of attainment gaps (although Tables 10.8 to 10.10 and 10.18 to 10.20 go beyond univariate analysis in considering the effect of SES at different levels of prior attainment). A shortcoming of univariate analyses is that, although they give a wealth of information, they can sometimes be misleading as they do not consider possible correlations between variables and variations in underlying characteristics of the groups that may explain observed or hidden differences. Multivariate analyses, which allow the effect of a variable to be examined while holding other variables constant, provide more nuanced results.

To this end, we carried out linear mixed effects modelling of 3 years' data. The analysis aimed to model the relationship between a numeric grade, on the one hand, and background information about the student that the entry belonged to and the year and subject of the entry, on the other. Centres and students within centres were treated as random effects with random intercepts in the model.

The fixed effects included were:

- Gender: male, female (reference category), unknown
- Prior attainment: low, mid (reference category), high, unknown
- Ethnicity: AOEG, ASIA, BLAC, CHIN, MIXD, UNCL, WHIT (reference category), unknown
- Language: ENG (reference category), OTH, UNCL, unknown
- SEN: NON (reference category), SNS, SS, UNCL, unknown
- FSM: NO (reference category), YES, unknown
- SES: low, mid (reference category), high, unknown
- Subject:[126]
    - For A level analysis: a range of subjects, with psychology as the reference category
    - For AS level analysis: a range of subjects, with mathematics as the reference category
- Year: 2018, 2019 (reference category), 2020
- Interactions: Gender*Year; Prior attainment*Year; Ethnicity*Year; Language*Year; SEN*Year; FSM*Year; SES*Year; Subject*Year

It is important to note that because our prior attainment measure was based on GCSE performance and there were correlations between most variables in the model and prior attainment (as can be verified by examining the prior attainment means at different levels of each variable in Tables 10.2 to 10.7 and Tables 10.12 to 10.17), the effects of many variables on A or AS level outcome were likely to be wrapped up in their effects on GCSE attainment. As the model quantifies the effect of each variable after controlling for prior attainment, among other variables, the effects relate to changes between groups that would have taken place between students taking their GCSEs and their A or AS level rather than the effects which may be introduced across an entire school career.

10.4.1 A level

Estimates of the parameters of the model for A level are presented in Table 10.22, except for the estimates of the parameters relating to the *Subject* variable. The *Subject* main effects and interactions with *Year* tell us about inter-subject comparability and any change in inter-subject comparability between years. They are omitted from Table 10.22 because they do not address equality issues related to the student background variables given in Section 10.2.

The left hand third of the table presents results of the main effects, which tell us about the relationships between background variables and grade in 2019. According to the model, a 'modal' student taking A level psychology in 2019 who was in the reference category of every background variable (that is, white, female, mid SES, not FSM eligible, with English as major language, no SEN and a medium level of prior

[126] For the AS level analysis, all data in all subjects analysed in the univariate analyses were included. For the A level analysis, because of the limitation of computing power, only the ten most popular subjects were included in the modelling, with over one million entries over three years, representing 78% of the data used in the univariate analyses. The subject with the most entries was taken as the reference category of the Subject variable in the respective analysis.

attainment) would be awarded grade of 3.33 (somewhere between grades B and C), shown by the intercept estimate of the model. The regression coefficient of each contrast indicates how different the grade received by a student (from the same centre) differing from the modal student by only one attribute would be from the modal student's grade. For example, a student with a high level of prior attainment who was from the same centre as the modal student and who shared with the modal student the attributes of being white, female, mid SES, not FSM eligible and having English as major language and no SEN would receive the grade 3.33+1.19=4.52 (somewhere between grade A and B) in A level psychology in 2019. The regression coefficient of each contrast indexes the magnitude of the relevant attainment gap in 2019 after controlling for other variables. The t value was obtained by dividing a regression coefficient by its standard error. The t value can be compared to the critical value of 1.96 (for $p<.05$) or 2.54 (for $p<.01$) to determine the statistical significance of the attainment gap indicated by the regression coefficient. However, this practice is not universally accepted by experts on mixed effects modelling (see, for example, Bates, 2006),[127] and there are questions over the value of conventional tests of statistical significance in analyses of large administrative datasets (see, for example, Connelly et al., 2016).[128]

To provide a measure of the practical (as opposed to statistical) significance or substantive importance of each contrast, it is possible to consider the estimate of the regression coefficient. Alternatively, a standardised effect size, namely Cohen's d adapted for the multilevel framework, was calculated (see, for example, Hedges, 2007; Westfall, 2016).[129,130] Cohen's original classification of effect sizes of 0.2/0.5/0.8 as small/medium/large is now widely recognised as not applicable to every context. In education, given the difficulty in raising academic achievement, it is recognised that measures that have effect sizes smaller than Cohen's small effect are still of educational significance (see Coe, 2002 and Hill et al., 2008 for discussions on interpreting effect sizes in education research).[131,132] For the present purpose, we set the threshold at the highly cautious level |0.1| for highlighting an effect as of substantive importance, this means marginal effects may be identified.

The first part of the table tells us that after controlling for other variables, the effects of most variables in 2019 were not of a magnitude of substantive importance, and the ones that were are the differences relating to prior attainment, the higher

127 Bates, D. (2006). lmer, p-values and all that. Internet post accessed on 5 August 2020 at https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

128 Connelly, R., Playford, C.J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research, 59*, 1-12.

129 Hedges, L.V. (2007). Effect sizes in cluster-randomized design. *Journal of Educational and Behavioral Statistics, 32*, 341-370.

130 Westfall, J. (2016). Five different "Cohen's d" statistics for within-subject designs. Blogpost accessed on 5 August 2020 at http://jakewestfall.org/blog/index.php/category/effect-size/

131 Coe, R. (2002). It's the effect size, stupid. What effect size is and why it is important. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England, 12-14 September 2002.

132 Hill, C.J., Bloom, H.S., Black, A.R., & Lipsey, M.W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2,* 172–177.

attainment of male students relative to female students with comparable background characteristics, and the higher attainment of Chinese students relative to white students with comparable background characteristics. It should be remembered that when we conclude from the standardised effect size the effect of a variable (that is, being X [versus Y]) in the model, we make a statement about the substantive importance or practical significance of the difference that being X [versus Y] *between GCSE and A level* makes to A level outcome, not the substantive importance or practical significance of the difference that being X [versus Y] in the entire school career makes to A level outcome. Saying that being X [versus Y] between GCSE and A level makes little substantively important or practically significant difference to one's A level grade does not amount to saying that being X [versus Y] in one's entire school career makes little substantively important or practically significant difference to one's A level grade.

The middle part of Table 10.22 presents results of the interactions with the Year: 2018 variable, which tell us about the difference between effects in 2019 (presented in the first part of the table) and effects in 2018. We saw above that the modal student would receive grade 3.33 in 2019; the first row in the middle part says that according to the model, in 2018 a modal student (who must be a different person to, but shared all attributes with, the 2019 modal student) would receive grade 3.33+0.02=3.35. And while the difference in grade between two students who differed only in their high versus medium level of prior attainment was 1.19 in 2019, the same difference was 0.03 grades lower, that is, 1.19-0.03=1.16 grades in 2018. So, in the middle part, the regression coefficient of each contrast indexes the magnitude of the change between 2018 and 2019 in the relevant attainment gap after controlling for other variables, and the standardised effect size indexes the substantive importance of the magnitude of the change.

As can be seen from the highlighting in the middle part of the table, only one change between 2018 and 2019 – those relating to students of unknown gender – is marked as of substantive importance. There were very few students of unknown gender (see Table 10.12).

The third section of the table is the most informative in relation to whether any existing attainment gap has been exacerbated in 2020. The interpretation of the statistics is as in the middle part of the table, so the regression coefficient of each contrast indexes the magnitude of the change in gap between 2019 and 2020 after controlling for other variables. As can be seen from the highlighting in the third part of Table 10.22, the only change between 2019 and 2020 marked as being of substantive importance is the disappearance of the higher attainment of male students relative to female students. The model suggests that male students outperformed female students with comparable background characteristics by 0.11 grade in 2019, 0.11+0.05=0.16 grade in 2018 and 0.11-0.12=-0.01 grade in 2020. The change between 2019 and 2020 marked as being of substantive importance can be seen as continuing a change which has already occurred between 2018 and 2019 and does not appear to have resulted from the 2020 awarding process. We saw in the univariate analysis of A level a growth in gap between the grades of female students compared to males between 2018 and 2020. The multivariate modelling suggests that the change can better be understood as a decline of male students' attainment relative to that of female students with comparable background characteristics. Importantly, this change has already occurred between 2018 and 2019, and so does not appear to be caused by the 2020 awarding process.

*Table 10.22 Parameter estimates of linear mixed effect model of effects of student background variables on A Level grades (Subject main effects and *Subject interactions omitted)*

	Effects in 2019				Effects in 2018: interaction between each term and Year: 2018				Effects in 2020: interaction between each term and Year: 2020			
	Regression coefficient	**Std. Error**	**t value**	**Standardised effect size**	**Regression coefficient**	**Std. Error**	**t value**	**Standardised effect size**	**Regression coefficient**	**Std. Error**	**t value**	**Standardised effect size**
(Intercept)	3.33	0.01	333.75	2.89	0.02	0.01	2.39	0.02	0.03	0.01	3.43	0.03
Prior attainment: High	1.19	0.01	213.83	1.04	-0.03	0.01	-3.37	-0.02	0.04	0.01	4.89	0.03
Prior attainment: Low	-0.90	0.01	-153.04	-0.78	0.03	0.01	4.00	0.03	0.01	0.01	1.47	0.01
Prior attainment: Unknown	0.37	0.01	36.68	0.32	0.01	0.01	0.49	0.01	-0.03	0.01	-2.09	-0.02
Gender: Male	0.11	0.00	23.10	0.10	0.05	0.01	6.82	0.04	-0.12	0.01	-17.75	-0.10
Gender: Unknown	-0.11	0.18	-0.60	-0.09	1.24	0.72	1.72	1.07				
Ethnicity: AOEG	-0.01	0.02	-0.47	-0.01	0.03	0.03	0.93	0.02	0.00	0.03	0.02	0.00
Ethnicity: ASIA	-0.04	0.01	-4.61	-0.04	0.01	0.01	0.95	0.01	-0.03	0.01	-2.49	-0.03
Ethnicity: BLAC	-0.09	0.01	-7.27	-0.08	0.00	0.02	-0.06	0.00	-0.05	0.02	-2.90	-0.04
Ethnicity: CHIN	0.18	0.03	6.29	0.16	-0.06	0.04	-1.35	-0.05	0.01	0.04	0.20	0.01
Ethnicity: MIXD	-0.03	0.01	-2.38	-0.02	0.01	0.02	0.51	0.01	-0.02	0.02	-1.30	-0.02
Ethnicity: UNCL	-0.03	0.02	-1.50	-0.03	-0.03	0.03	-0.88	-0.02	-0.04	0.03	-1.28	-0.03
Ethnicity: Unknown	0.02	0.06	0.27	0.01	-0.03	0.08	-0.41	-0.03	0.01	0.08	0.09	0.01
FSM: Yes	-0.04	0.01	-3.33	-0.03	-0.04	0.02	-2.29	-0.03	-0.04	0.02	-2.39	-0.03

Language: OTH		-0.03	0.01	-3.44	-0.03	-0.01	0.01	-0.94	-0.01	0.03	0.01	2.11	0.02
Language: UNCL		-0.04	0.04	-0.93	-0.03	-0.08	0.06	-1.33	-0.07	0.01	0.05	0.24	0.01
SEN: SNS		0.02	0.01	1.45	0.02	-0.05	0.02	-2.99	-0.05	-0.09	0.02	-5.01	-0.07
SEN: SS		0.01	0.02	0.61	0.01	-0.01	0.03	-0.38	-0.01	0.02	0.03	0.60	0.02
SES: High		0.05	0.01	7.55	0.04	0.00	0.01	-0.51	0.00	-0.03	0.01	-3.39	-0.02
SES: Low		-0.05	0.01	-7.45	-0.04	0.00	0.01	-0.09	0.00	0.03	0.01	3.86	0.03
SES: Unknown		-0.05	0.06	-0.81	-0.04	0.03	0.08	0.39	0.03	-0.02	0.08	-0.30	-0.02

Random effects:
Student variance: 0.693
Centre variance: 0.133
Residual variance: 0.505

10.4.2 AS level

Estimates of the parameters of the model for AS level are presented in Table 10.23, except for the estimates of the parameters relating to the Subject variable, which are omitted because they do not address equality issues related to the student background variables given in Section 10.2.

As can be seen in Table 10.23, after controlling for other variables, the effects of most variables in 2019 were not of a magnitude of substantive importance, and the ones that were are the differences relating to prior attainment, the under-attainment of students of unknown gender relative to male students, the under-attainment of black students relative to white students, the over-attainment of Chinese students relative to white students, and the under-attainment of students with an unclassified major language relative to those with English as their major language. Again it should be remembered that when we conclude from the standardised effect size of the effect of a variable (that is, being X [versus Y]) in the model, we make a statement about the substantive importance or practical significance of the difference that being X [versus Y] *between GCSE and AS level* makes to AS level outcome, not the substantive importance or practical significance of the difference that being X [versus Y] in the entire school career makes to AS level outcome. Saying that being X [versus Y] between GCSE and AS level makes little substantively important or practically significant difference to one's AS level grade does not amount to saying that being X [versus Y] in one's entire school career makes little substantively important or practically significant difference to one's AS level grade.

As can be seen from the highlights in the middle part, three changes between 2018 and 2019 are marked as of substantive importance following the threshold of |0.1|: narrowing of the lower attainment of female students relative to male students from 2018 to 2019, the emergence in 2019 of the lower attainment of students with an unclassified major language relative to those with English as their major language (which did not exist in 2018), narrowing of the higher attainment of Statemented students with SEN relative to students without SEN from 2018 to 2019.

The third part of Table 10.23 presents results of the interactions with the Year: 2020 variable, which tell us about the difference between effects in 2019 (presented in the first part of the table) and effects in 2020. As can be seen from the highlights, two changes between 2019 and 2020 are marked as of substantive importance following the threshold of |0.1|.

One relates to the higher attainment of students with an unclassified major language relative to those with English as their major language after controlling for other variables. We saw above that this attainment gap emerged in 2019. According to the model, this attainment gap has disappeared in 2020.

The other change whose magnitude is marked as being of substantive importance relates to the change from male students outperforming female students with comparable background characteristics in 2019 to female students outperforming male students with comparable background characteristics in 2020. The model suggests that male students outperformed female students with comparable background characteristics by 0.06 grade in 2019, 0.06+0.14=0.2 grade in 2018 and 0.06-0.13=-0.07 grade in 2020. The change between 2019 and 2020 marked as being of substantive importance could be seen as continuing a change which has occurred between 2018 and 2019.

Table 10.23. Parameter estimates of linear mixed effect model of effects of student background variables on AS Level grades

	Effects in 2019				Effects in 2018: interaction between each term and Year: 2018				Effects in 2020: interaction between each term and Year: 2020			
	Regression coefficient	Std. Error	t value	Standardised effect size	Regression coefficient	Std. Error	t value	Standardised effect size	Regression coefficient	Std. Error	t value	Standardised effect size
(Intercept)	2.34	0.03	74.58	1.74	-0.10	0.03	-3.19	-0.08	-0.01	0.03	-0.33	-0.01
Prior attainment: High	1.38	0.02	75.84	1.03	-0.07	0.03	-2.68	-0.05	0.00	0.03	-0.12	0.00
Prior attainment: Low	-1.11	0.02	-61.30	-0.82	0.02	0.02	0.78	0.01	0.03	0.03	1.22	0.02
Prior attainment: Unknown	0.12	0.03	3.89	0.09	-0.04	0.04	-0.95	-0.03	0.08	0.05	1.72	0.06
Gender: Male	0.06	0.01	3.89	0.04	0.14	0.02	7.07	0.11	-0.13	0.02	-6.24	-0.10
Gender: Unknown	-0.33	0.82	-0.40	-0.25								
Ethnicity: AOEG	-0.03	0.06	-0.45	-0.02	-0.02	0.08	-0.23	-0.01	0.03	0.08	0.43	0.03
Ethnicity: ASIA	-0.07	0.03	-2.59	-0.05	0.01	0.04	0.28	0.01	-0.05	0.04	-1.27	-0.03
Ethnicity: BLAC	-0.14	0.03	-4.29	-0.10	0.06	0.04	1.33	0.04	-0.05	0.05	-1.11	-0.04
Ethnicity: CHIN	0.28	0.08	3.26	0.21	-0.08	0.12	-0.70	-0.06	-0.06	0.12	-0.46	-0.04
Ethnicity: MIXD	-0.01	0.03	-0.18	0.00	-0.06	0.05	-1.16	-0.04	0.00	0.05	0.02	0.00
Ethnicity: UNCL	-0.03	0.07	-0.47	-0.02	-0.03	0.09	-0.32	-0.02	0.02	0.10	0.26	0.02
Ethnicity: Unknown	-0.10	0.19	-0.53	-0.07	0.09	0.26	0.36	0.07	0.03	0.25	0.14	0.03
FSM: Yes	-0.03	0.03	-0.91	-0.02	-0.08	0.04	-1.82	-0.06	0.01	0.04	0.18	0.01
Language: OTH	-0.05	0.02	-1.94	-0.03	0.04	0.03	1.36	0.03	0.07	0.03	2.00	0.05
Language: UNCL	-0.22	0.11	-1.92	-0.16	0.25	0.16	1.50	0.18	0.19	0.16	1.15	0.14
SEN: SNS	-0.02	0.04	-0.57	-0.02	0.02	0.05	0.45	0.02	-0.09	0.05	-1.67	-0.07
SEN: SS	0.06	0.08	0.77	0.05	0.14	0.11	1.30	0.11	0.05	0.10	0.54	0.04
SES: High	0.04	0.02	1.85	0.03	-0.01	0.03	-0.38	-0.01	0.02	0.03	0.67	0.01
SES: Low	-0.03	0.02	-1.27	-0.02	-0.04	0.03	-1.35	-0.03	0.00	0.03	0.04	0.00
SES: Unknown	0.07	0.19	0.39	0.05	-0.10	0.26	-0.38	-0.07	-0.12	0.25	-0.47	-0.09

Random effects: Student variance: 0.668, Centre variance: 0.232, Residual variance: 0.903

10.5 Multivariate analysis: focus on specific subjects

The modelling exercise presented above was conducted by pulling together, separately for AS and A levels, data from a wide range of subjects and over the three-year period 2018-2020. Although this has the advantage of providing an overall picture, a subject level approach is also informative. The analysis was therefore conducted for a small sample of A level subjects individually and separately for each year

This analysis focuses on two A level subjects: maths and French. Mathematics was chosen as it is a large entry subject, outside of the scope of those included in the analysis presented above. As above, the aim was to model the relationship between students' attainment and their socio-economic and demographic background, once prior attainment is accounted for. We used multilevel regression models to account for the hierarchical structure of the data (that is, students clustered within schools). To allow for as much flexibility as possible, rather than estimating only one model across the three years, three different regression models were fitted, one per each year. The figures presented below cannot, therefore, be immediately compared with those presented in previous sections. Similar to the results shown above, results for 2019 and 2018 can be used to interpret results for 2020 by providing a benchmark of what the gap looked like in recent years. More specifically, this approach allows us to evaluate how the attainment gap for students with protected characteristics has changed over time and therefore to evaluate whether in 2020 it was wider than in previous years.

We measured the attainment both as a point score (by converting grades into numbers, ie A*=6, A=5,…) and as the probability of attaining grade A and C (or above). Where probabilities were considered as dependant variables, both linear probability models and logistic regressions were estimated. Results of the two approaches were very similar, therefore for ease of interpretation only the results of the linear probability models are reported below.

Table 10.24 refers to the results for A level maths. The first set of columns in the table report the results for the attainment measured as a point score. Given the conversion used to translate letter grades into point scores, 1 unit can be interpreted as 1 grade. So, for example, when in 2019 the estimate of the coefficient for male students is 0.51, it means that boys tend to perform half a grade better than girls, once other factors are controlled for. In 2020 the estimate of this coefficient is 0.34, which can be interpreted as a difference of one third of a grade between boys and girls. This suggests that the gender gap in 2020 is slightly narrower than it was the year before. However, the gap was 0.35 in 2018 so the 2020 gap almost exactly matches the gap in 2018. Results of the linear probability models for the probability of attaining a grade A and C (or above) confirm these results. In both cases the gap in the probability of attaining the grade reduced from 2019 to 2020, from 15 to 10 percentage points at grade A and from 8 to 5 percentage points at grade C. In each case the gap in 2020 was closer to the gap seen in 2018. This suggests the differences seen in 2020 are normal fluctuations and not evidence of bias in the process of awarding grades. However, it should be noted that the cohort of students who entered A level maths in 2018 was small and potentially selective.

The second protected characteristic of interest is Special Education Needs status. From Table 10.24 it is apparent that any gap in attainment for students with special education needs almost vanished in 2020. Opposite results were found for students for which English is not their first language. In this case, in fact, the gap seems to have widened in 2020 for this subject. The size of the gap, however, appears to be very small. Once attainment is controlled for, EAL students seem to be performing better than their English native speaker counterparts by only 2 percentage points at grade A and by only 1 percentage point at grade C. This is comparable to the gap in 2018.

For ethnicity the findings vary across different ethnic groups although the size of the fluctuations are similar to those seen between 2018 and 2019 and so do not suggest that the process of awarding grades has introduced bias. The largest gap was found for Chinese pupils who tend to outperform White pupils with similar characteristics by, on average, slightly more than one quarter of a grade (0.27) in 2020, slightly higher than in 2019 (0.23). This seems to suggest a widening of the Chinese-White pupils gap, however, this figure conceals a more complicated picture, with the gap getting wider at grade A, but narrower at grade C. The difference between 2019 and 2020 is almost negligible. Similar findings were found for Black pupils. The gap in 2020 can be quantified as 0.12 of a grade, closely aligned with the gap found in 2019 (0.11). A similar pattern was found for the other ethnic groups.

The last protected characteristic considered in this analysis was socio-economic status, measured by the IDACI score[133]. The gap between groups was very small and tended to narrow in 2020. Again, the size of the fluctuations was similar to that seen between 2018 and 2019 and so do not suggest any bias in the grading process in 2020.

Results for A level French displayed in Table 10.25 are very similar to those found for mathematics: once controlled for prior attainment, gaps in grades achieved in 2020 among different groups of students are usually very small and tend to be not wider than in recent years. For French one exception is gender: boys appear 2 percentage points more likely than girls to attain a grade A in 2020 than in 2019. However, the gap in the probability of attaining a grade C remain stable over time. Moreover, there was a similar change in the gender gap between 2018 and 2019. The change this year is similar to the size of changes that seen when exams are taken.

[133] Free School Meal eligibility could also be considered in addition or as an alternative to IDACI. Here IDACI was considered instead for the high correlation with FSM eligibility.

Table 10.24. Regression estimates for the role of socio-economic and demographic characteristics on attainment in 2018, 2019 and 2020 - A level mathematics

	Point scores (1 unit = 1 grade)			**Probability of attaining a grade**					
				Grade A (or above)			**Grade C (or above)**		
	2018	**2019**	**2020**	**2018**	**2019**	**2020**	**2018**	**2019**	**2020**
(Intercept)	-2.52	-2.93	-2.93	-0.41	-1.34	-1.42	-0.71	-0.58	-0.50
	(0.39)	(0.08)	(0.06)	(0.07)	(0.02)	(0.02)	(0.12)	(0.02)	(0.02)
Gender: Male (vs Female)	0.35	0.51	0.34	0.03	0.15	0.10	0.06	0.08	0.05
	(0.11)	(0.01)	(0.01)	(0.02)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)
SEN: Yes (vs none)	-0.23	0.15	0.00	0.01	0.04	-0.00	-0.05	0.03	-0.00
	(0.21)	(0.03)	(0.02)	(0.04)	(0.01)	(0.01)	(0.06)	(0.01)	(0.01)
EAL: Yes (vs none)	0.08	0.01	0.06	0.02	0.00	0.02	0.02	-0.00	0.01
	(0.14)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.04)	(0.01)	(0.00)
Ethnicity: AOEG (vs White)	-0.10	-0.05	0.04	-0.05	-0.01	0.01	-0.08	-0.01	0.03
	(0.33)	(0.04)	(0.04)	(0.06)	(0.01)	(0.01)	(0.10)	(0.01)	(0.01)
Ethnicity: Asian (vs White)	0.03	-0.11	-0.06	-0.03	-0.02	-0.01	-0.02	-0.02	-0.01
	(0.15)	(0.02)	(0.02)	(0.03)	(0.01)	(0.01)	(0.05)	(0.01)	(0.01)
Ethnicity: Black (vs White)	-0.03	-0.11	-0.12	-0.03	-0.03	-0.03	-0.05	-0.01	-0.02
	(0.19)	(0.03)	(0.02)	(0.03)	(0.01)	(0.01)	(0.06)	(0.01)	(0.01)
Ethnicity: Chinese (vs White)	0.99	0.23	0.27	0.20	0.06	0.09	0.25	0.05	0.03
	(0.62)	(0.05)	(0.05)	(0.11)	(0.02)	(0.02)	(0.19)	(0.01)	(0.01)
Ethnicity: Mixed (vs White)	0.29	-0.02	-0.04	-0.01	-0.00	-0.01	0.03	0.00	0.01
	(0.22)	(0.03)	(0.02)	(0.04)	(0.01)	(0.01)	(0.07)	(0.01)	(0.01)
Deprivation: High (vs low)	0.01	-0.09	-0.05	0.02	-0.02	-0.02	0.02	-0.02	-0.01
	(0.13)	(0.02)	(0.01)	(0.02)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)
Deprivation: Medium (vs low)	0.06	-0.06	-0.03	-0.01	-0.02	-0.01	0.04	-0.01	-0.01
	(0.13)	-2.93	(0.01)	(0.02)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)
N (students)	873	56216	60161	873	56216	60161	873	56216	60161
N (CentreNo)	411	2523	2466	411	2523	2466	411	2523	2466
R2 (fixed)	0.15	0.28	0.41	0.08	0.25	0.32	0.10	0.16	0.22
R2 (total)	0.29	0.37	0.44	0.11	0.29	0.33	0.23	0.23	0.24

Note: Prior attainment and centre-level variables not reported in the table. Standard errors reported in brackets.

Table 10.25. Regression estimates for the role of socio-economic and demographic characteristics on attainment in 2018, 2019 and 2020 - A level French

	Point scores (1 unit = 1 grade)			**Probability of attaining a grade**					
				Grade A (or above)			**Grade C (or above)**		
	2018	**2019**	**2020**	**2018**	**2019**	**2020**	**2018**	**2019**	**2020**
(Intercept)	1.39	0.94	-0.82	-0.80	-0.81	-1.31	0.33	0.21	0.27
	(0.16)	(0.15)	(0.09)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Gender: Male (vs Female)	-0.01	0.08	0.14	0.02	0.04	0.06	0.00	0.01	0.01
	(0.05)	(0.05)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SEN: Yes (vs none)	0.23	0.09	-0.00	0.10	0.02	0.01	0.04	0.02	-0.02
	(0.10)	(0.10)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
EAL: Yes (vs none)	0.13	0.17	0.34	0.10	0.08	0.13	0.01	0.04	0.03
	(0.07)	(0.07)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Ethnicity: AOEG (vs White)	0.05	0.49	0.05	-0.11	0.06	0.00	0.03	0.12	0.06
	(0.18)	(0.17)	(0.12)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
Ethnicity: Asian (vs White)	0.06	-0.03	-0.13	-0.05	-0.03	-0.08	0.01	0.01	0.01
	(0.11)	(0.10)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Ethnicity: Black (vs White)	0.02	0.04	0.01	0.01	-0.02	0.01	0.01	0.03	0.00
	(0.11)	(0.10)	(0.06)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)
Ethnicity: Chinese (vs White)	-0.11	0.34	0.05	-0.04	0.07	0.13	0.04	0.00	0.04
	(0.30)	(0.37)	(0.20)	(0.08)	(0.10)	(0.09)	(0.08)	(0.10)	(0.06)
Ethnicity: Mixed (vs White)	0.06	0.11	0.00	0.02	0.04	-0.01	-0.01	0.02	-0.00
	(0.09)	(0.09)	(0.05)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Deprivation: High (vs low)	0.05	-0.01	-0.00	-0.01	0.01	-0.01	0.01	-0.02	-0.01
	(0.06)	(0.06)	(0.04)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)
Deprivation: Medium (vs low)	0.01	0.07	-0.01	-0.01	0.00	0.01	-0.01	0.03	-0.00
	(0.05)	(0.05)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
N	5154	5056	5081	5154	5056	5081	5154	5056	5081
N (CentreNo)	1378	1290	1263	1378	1290	1263	1378	1290	1263
R2 (fixed)	0.04	0.06	0.44	0.13	0.14	0.35	0.02	0.04	0.14
R2 (total)	0.14	0.13	0.47	0.16	0.18	0.36	0.13	0.11	0.14

Note: Prior attainment and centre-level variables not reported in the table. Standard errors reported in brackets.

10.6 Conclusion

The analyses conducted show no evidence that this year’s process of awarding grades has introduced bias. Changes in outcomes for students with different protected characteristics and from different socio-economic backgrounds are similar to those seen between 2018 and 2019.

11 Closing remarks

The 2020 exam series has been exceptional. In the absence of formal assessment evidence, we have sought to identify the fairest approach to awarding grades that will enable students to progress along their chosen route.

The standardisation model that was put in place aimed to protect against advantage or disadvantage to students at centres who took different approaches to producing their CAGs. An important aspect of the design of the model was to ensure that the approach taken in one centre did not impact negatively on the outcomes for students in another.

Part of this focus on fairness was also to make sure that the grades are credible and broadly in line with those from previous years. This reflects fairness to students in past and future years, but also those in the current cohort as it delivers results that have a value equivalent to those awarded in other years.

To achieve this, the approach developed for use this year sought to maintain overall outcomes, as far as possible, whilst incorporating design decisions which acted in the favour of candidates. On the evidence of the overall outcomes, this appears to have been broadly achieved with outcomes that are higher than in previous years, but which are meaningful reflections of what students may have achieved.

Overall, the results delivered this year will have met the aim of enabling large numbers of students to move on to the next stages of their lives. Further, there is no evidence of systemic assessment bias affecting the outcomes of students with particular protected characteristics or from different socio-economic backgrounds.

This interim report has provided a description of the standardisation approach that was implemented by exam boards, the steps taken to develop that approach and an overview of the grades that have resulted from its use for AS and A level qualifications.

On GCSE results day, we will publish relevant analyses of the those results and any other useful contextual information.

A final report will be published later in the year when we have completed our evaluation of this summer’s results. This will include reporting on work such as that introduced in Annex R, which explores the experiences of teachers in producing their CAGs and rank orders this summer.

Annex A: In-scope subject list

GCSE (Full course)

Subject	Phase of reform
English language	Phase 1
English literature	
Mathematics	
Classical Greek	Phase 2
Latin	
Art and design	
Biology	
Chemistry	
Citizenship studies	
Combined science (double award)	
Computer science	
Dance	
Drama	
Food preparation and nutrition	
Geography	
History	
French	
German	
Spanish	
Music	
Physics	
Physical education	
Religious studies	
Ancient history	Phase 3
Astronomy	
Business	
Classical civilisation	

Subject	Phase of reform
Design and technology	Phase 3
Economics	
Electronics	
Engineering	
Film studies	
Geology	
Media studies	
Arabic	
Bengali	
Chinese	
Italian	
Japanese	
Modern Greek	
Modern Hebrew	
Panjabi	
Polish	
Russian	
Urdu	
Psychology	
Sociology	
Statistics	
Biblical Hebrew	Phase 4
Gujarati	
Persian	
Portuguese	
Turkish	

GCSE (Short course)

Subjects	Phase of reform
Physical education	Phase 2
Religious studies	

AS

Subject	Phase of reform
Art and design	Phase 1
Biology	
Business	
Chemistry	
Computer science	
Economics	
English language	
English language and literature	
English literature	
History	
Physics	
Psychology	
Sociology	
Classical Greek	Phase 2
Latin	
Dance	
Drama and theatre	
Geography	
French	
German	
Spanish	
Music	

Subject	Phase of reform
Physical education	Phase 2
Religious studies	
Accounting	Phase 3
Ancient history	
Classical civilisation	
Design and technology	
Electronics	
Environmental science	
Film studies	
Further mathematics	
Geology	
Government and politics	
Law	
Mathematics	
Media studies	
Chinese	
Italian	
Russian	
Music technology	
Philosophy	
Statistics	

A-level

Subject	Phase of reform
Art and design	Phase 1
Biology	
Business	
Chemistry	
Computer science	
Economics	
English language	

Subject	Phase of reform
Classical Greek	Phase 2
Latin	
Dance	
Drama and theatre	
Geography	
French	
German	

English language and literature	
English literature	
History	
Physics	
Psychology	
Sociology	
Classical civilisation	Phase 3
Design and technology	
Electronics	
Environmental science	
Film studies	
Further mathematics	
Geology	
Government and politics	
History of art	
Law	
Mathematics	
Media studies	
Chinese	
Italian	
Russian	
Music technology	

Spanish	
Music	
Physical education	
Religious studies	
Accounting	Phase 3
Ancient history	
Philosophy	Phase 3
Statistics	
Biblical Hebrew	Phase 4
Arabic	
Bengali	
Gujarati	
Greek	
Japanese	
Modern Hebrew	
Panjabi	
Persian	
Polish	
Portuguese	
Turkish	
Urdu	

Other

Subjects	Phase of reform
EPQ	N/A
AEA in mathematics	

Annex B:[134] Post-results changes

Post-results changes – A level

Change in cumulative % outcomes following reviews of marking and moderation in 2019 by subject and grade

	Subject level grade						
Subject	A*	A	B	C	D	E	U
Accounting	0.7	0.7	0.7	0.9	0.1	0.1	0.0
Ancient history	0.3	0.3	0.3	0.0	0.0	0.0	0.0
Art	0.2	0.1	0.1	0.1	0.0	0.0	0.0
Art: 3D studies	0.0	0.0	0.1	0.2	0.0	0.0	0.0
Art: Critical studies	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: Fine art	0.3	0.2	0.2	0.2	0.0	0.0	0.0
Art: Graphics	0.0	0.1	0.0	0.1	0.0	0.0	0.0
Art: History of art	0.5	-0.6	-0.1	0.1	0.0	0.0	0.0
Art: Photography	0.3	0.2	0.5	0.2	0.0	0.0	0.0
Art: Textiles	0.7	0.2	0.5	0.3	0.1	0.1	0.0
Biology	0.6	0.4	0.6	0.4	0.1	0.1	0.0
Business studies	0.6	0.5	0.8	0.5	0.1	0.0	0.0
Chemistry	0.4	0.1	0.3	0.2	0.1	0.0	0.0
Chinese	0.5	0.5	0.0	0.1	0.0	0.0	0.0
Classical civilisation	1.1	0.8	1.1	0.6	0.0	0.0	0.0
Classical Greek	0.5	0.0	0.0	0.0	0.0	0.0	0.0
Computing	0.6	0.2	0.4	0.4	0.2	0.1	0.0
Dance	1.0	1.0	1.0	0.4	0.1	0.0	0.0
Design and technology	0.5	0.4	0.5	0.4	0.1	0.0	0.0
Drama	0.4	0.3	0.5	0.3	0.1	0.0	0.0
Economics	0.9	0.4	0.6	0.3	0.1	0.0	0.0
Electronics	0.0	0.0	0.0	0.0	0.2	0.0	0.0
English language	0.3	0.2	0.3	0.3	0.0	0.0	0.0
English lang & lit	0.7	0.5	0.9	0.5	0.0	0.0	0.0
English literature	1.1	0.7	0.8	0.3	0.1	0.0	0.0
Environmental studies	0.3	0.1	0.8	0.7	0.2	0.1	0.0
Film studies	0.8	0.5	0.9	0.3	0.0	0.0	0.0
French	0.5	0.2	0.3	0.1	0.0	0.0	0.0
Further mathematics	0.3	-0.2	0.2	0.1	0.0	0.0	0.0
Geography	0.9	0.6	0.6	0.3	0.1	0.0	0.0
Geology	0.3	0.4	0.7	0.3	0.1	0.2	0.0
German	0.5	0.0	0.0	0.2	0.0	0.0	0.0
History	0.8	0.6	0.6	0.3	0.1	0.0	0.0
Italian	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Latin	1.6	0.7	0.7	0.1	0.0	0.0	0.0

134 Original outcomes based on data submitted to Ofqual by exam boards approximately one week before results were issued (any missing data is likely to be missing at random). Post-results outcomes based on reviews of marking and moderation data submitted to Ofqual by exam boards.

Law	0.3	0.2	0.3	0.1	0.0	0.0	0.0
Mathematics	0.3	0.0	0.3	0.2	0.1	0.0	0.0
Media studies	0.6	0.5	0.5	0.2	0.0	0.0	0.0
Music	0.9	0.6	1.2	0.8	0.1	0.0	0.0
Philosophy	0.3	0.1	0.2	0.2	0.0	0.1	0.0
Physical education	0.9	0.6	0.9	0.7	0.3	0.1	0.0
Physics	0.5	0.1	0.4	0.2	0.1	0.1	0.0
Political studies	0.9	0.5	0.4	0.1	0.0	0.0	0.0
Psychology	0.5	0.3	0.4	0.2	0.1	0.0	0.0
Religious studies	0.5	0.3	0.4	0.2	0.0	0.0	0.0
Russian	0.0	-0.7	0.0	0.0	0.0	0.0	0.0
Sociology	0.5	0.3	0.6	0.3	0.1	0.0	0.0
Spanish	0.8	0.2	0.7	0.3	0.1	0.0	0.0
Statistics	0.2	-0.1	0.6	0.5	0.2	0.0	0.0

Post-results changes – AS

Change in cumulative % outcomes following reviews of marking and moderation in 2019 by subject and grade

	Subject level grade					
Subject	A	B	C	D	E	U
Accounting	0.6	0.0	0.6	0.6	0.0	0.0
Ancient history	0.0	0.0	0.0	0.0	0.0	0.0
Art	0.0	0.0	0.0	0.0	0.0	0.0
Art: 3D studies	0.0	0.0	0.0	0.0	0.0	0.0
Art: Critical studies	0.0	0.0	0.0	0.0	0.0	0.0
Art: Fine art	0.0	0.0	0.0	0.0	0.1	0.0
Art: Graphics	0.0	0.0	0.0	0.0	0.0	0.0
Art: Photography	0.0	0.0	0.1	0.1	0.1	0.0
Art: Textiles	0.0	0.0	0.0	0.0	0.0	0.0
Biology	0.2	0.1	0.1	0.1	0.1	0.0
Business studies	0.2	0.1	0.2	0.2	0.1	0.0
Chemistry	0.2	0.1	0.0	0.0	0.0	0.0
Chinese	0.0	0.0	0.0	0.0	0.0	0.0
Classical civilisation	0.0	0.0	0.0	0.3	0.0	0.0
Classical Greek	0.0	0.0	0.0	0.0	0.0	0.0
Computing	0.2	0.2	0.2	0.2	0.2	0.0
Dance	1.1	0.0	0.0	0.0	0.0	0.0
Design and technology	0.0	0.1	0.1	0.0	0.0	0.0
Drama	0.0	0.0	0.0	0.0	0.0	0.0
Economics	0.3	0.1	0.2	0.1	0.0	0.0
Electronics	0.0	0.0	0.0	0.0	0.5	0.0
English language	0.1	0.3	0.0	0.0	0.1	0.0
English lang & lit	0.1	0.2	0.0	0.0	0.0	0.0
English literature	0.2	0.3	0.2	0.1	0.0	0.0
Environmental studies	0.3	0.5	1.1	1.6	0.4	0.0
Film studies	0.1	0.3	0.2	0.0	0.0	0.0

French	0.0	0.0	0.1	0.2	0.0	0.0
Further mathematics	0.3	0.2	0.0	0.0	0.0	0.0
Geography	0.3	0.2	0.3	0.1	0.1	0.0
Geology	0.0	0.0	0.5	0.0	0.5	0.0
German	0.2	0.2	0.0	0.0	0.0	0.0
History	0.0	0.1	0.1	0.0	0.1	0.0
Italian	0.0	0.0	0.0	0.0	0.0	0.0
Latin	1.8	0.9	0.0	0.0	0.0	0.0
Law	0.0	0.0	0.1	0.0	0.1	0.0
Mathematics	0.1	0.1	0.1	0.1	0.1	0.0
Media studies	0.1	0.0	0.1	0.0	0.0	0.0
Music	0.4	0.3	0.3	0.2	0.4	0.0
Philosophy	0.2	0.0	0.0	0.2	0.0	0.0
Physical education	0.3	0.3	0.4	0.2	0.3	0.0
Physics	0.2	0.2	0.1	0.2	0.1	0.0
Political studies	0.1	0.0	0.1	0.2	0.1	0.0
Psychology	0.2	0.2	0.1	0.1	0.1	0.0
Religious studies	0.3	0.3	0.2	0.0	0.0	0.0
Russian	0.0	0.0	0.0	0.0	0.0	0.0
Sociology	0.1	0.3	0.3	0.1	0.1	0.0
Spanish	0.7	0.3	0.3	0.2	0.1	0.0

Post-results changes – GCSE

Change in cumulative % outcomes following reviews of marking and moderation in 2019 by subject and grade

	Subject level grade									
Subject	9	8	7	6	5	4	3	2	1	U
Ancient history	0.8	0.9	0.3	0.4	0.3	0.2	0.1	0.0	0.0	0.0
Arabic	0.0	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0
Art	0.1	0.2	0.3	0.5	0.4	0.4	0.1	0.0	0.0	0.0
Art: 3D studies	0.1	0.2	0.3	0.2	0.1	0.5	0.1	0.0	0.0	0.0
Art: Critical studies	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: Fine art	0.1	0.2	0.3	0.4	0.3	0.2	0.1	0.0	0.0	0.0
Art: Graphics	0.2	0.3	0.3	0.4	0.3	0.2	0.0	0.1	0.0	0.0
Art: Photography	0.0	0.1	0.1	0.3	0.2	0.2	0.0	0.0	0.0	0.0
Art: Textiles	0.2	0.2	0.4	0.4	0.5	0.4	0.1	0.0	0.0	0.0
Astronomy	0.2	0.2	0.4	0.4	0.2	0.2	0.0	0.0	0.0	0.0
Bengali	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Biology	0.3	0.3	0.3	0.2	0.2	0.1	0.0	0.0	0.0	0.0
Business studies	0.2	0.3	0.4	0.4	0.3	0.3	0.1	0.0	0.0	0.0
Chemistry	0.3	0.3	0.3	0.2	0.2	0.1	0.0	0.0	0.0	0.0
Chinese	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Citizenship studies	0.2	0.2	0.2	0.3	0.2	0.2	0.1	0.0	0.0	0.0
Classical civilisation	0.7	0.9	1.0	0.8	0.4	0.3	0.0	0.0	0.0	0.0
Classical Greek	0.5	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Computing	0.2	0.3	0.3	0.3	0.3	0.3	0.1	0.0	0.0	0.0
Dance	0.1	0.2	0.3	0.3	0.2	0.2	0.1	0.0	0.0	0.0
Design and technology	0.1	0.3	0.3	0.2	0.2	0.2	0.1	0.0	0.0	0.0
Drama	0.1	0.2	0.3	0.2	0.2	0.1	0.0	0.0	0.0	0.0
Economics	0.6	0.8	0.8	0.7	0.5	0.4	0.1	0.0	0.0	0.0
Electronics	0.2	0.1	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0
Engineering	0.0	0.0	0.2	0.2	0.1	0.2	0.1	0.0	0.0	0.0

English language	0.1	0.1	0.3	0.3	0.3	0.5	0.0	0.0	0.0	0.0
English literature	0.1	0.1	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0
Film studies	0.1	0.1	0.3	0.2	0.2	0.2	0.0	0.0	0.0	0.0
Food prep & nutrition	0.1	0.0	0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.0
French	0.1	0.1	0.1	0.1	0.2	0.2	0.1	0.0	0.0	0.0
Geography	0.2	0.2	0.3	0.2	0.2	0.2	0.1	0.0	0.0	0.0
Geology	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
German	0.1	0.1	0.1	0.1	0.1	0.2	0.0	0.0	0.0	0.0
Greek	0.2	0.3	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
History	0.2	0.2	0.2	0.1	0.2	0.1	0.1	0.0	0.0	0.0
Italian	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Japanese	0.0	0.1	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0
Latin	1.0	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Mathematics	0.1	0.1	0.1	0.1	0.2	0.2	0.0	0.0	0.0	0.0
Media studies	0.0	0.1	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0
Modern Hebrew	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Music	0.3	0.3	0.4	0.3	0.2	0.3	0.0	0.0	0.0	0.0
Panjabi	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Physical education	0.2	0.4	0.6	0.5	0.5	0.5	0.1	0.0	0.0	0.0
Physics	0.3	0.3	0.3	0.2	0.2	0.1	0.0	0.0	0.0	0.0
Polish	0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0
Psychology	0.2	0.2	0.4	0.3	0.2	0.3	0.1	0.0	0.0	0.0
Religious studies	0.2	0.2	0.2	0.2	0.1	0.1	0.0	0.0	0.0	0.0
Russian	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Sociology	0.1	0.1	0.2	0.2	0.1	0.2	0.0	0.0	0.0	0.0
Spanish	0.1	0.2	0.2	0.2	0.2	0.2	0.1	0.0	0.0	0.0
Statistics	0.1	0.2	0.2	0.1	0.1	0.2	0.1	0.0	0.0	0.0
Urdu	0.4	0.4	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0

Annex C:[135] Partial Absentees

Partial Absentees – A level

Percentage of 18-year-olds in England in 2019 that were partially absent by subject[136] and grade

Subject	Entry	Subject level grade							
		A*	A	B	C	D	E	U	Total
Accounting	1860	0.0	0.0	0.0	0.0	0.0	0.0	4.3	0.2
Ancient history	590	0.0	0.0	0.0	0.0	0.0	0.0	22.2	0.3
Arabic	290	0.0	0.0	0.0	0.0	0.0	5.0	6.7	0.7
Art	4555	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
Art: 3D studies	1125	0.0	0.0	0.0	0.0	0.0	2.2	0.0	0.1
Art: Critical studies	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: Fine art	12185	0.0	0.0	0.0	0.0	0.2	0.3	2.8	0.0
Art: Graphics	4300	0.0	0.0	0.1	0.1	0.0	0.0	2.0	0.1
Art: History of art	390	0.0	0.0	0.0	0.0	2.2	0.0	0.0	0.3
Art: Photography	10230	0.0	0.0	0.0	0.0	0.0	0.5	2.6	0.0
Art: Textiles	2725	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0
Biology	56175	0.0	0.0	0.0	0.0	0.0	0.1	1.3	0.1
Business studies	26810	0.0	0.0	0.0	0.0	0.1	0.5	6.0	0.2
Chemistry	47865	0.0	0.0	0.0	0.0	0.0	0.2	1.3	0.1
Chinese	885	0.0	0.0	0.0	0.0	11.8	33.3	66.7	0.7
Classical civilisation	2565	0.0	0.0	0.0	0.0	0.0	0.0	6.7	0.0
Classical Greek	190	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Computing	9490	0.0	0.0	0.0	0.0	0.0	0.7	3.7	0.3
Dance	965	0.0	0.0	0.0	0.0	0.0	0.0	40.0	0.2
Design and technology	8670	0.0	0.0	0.0	0.0	0.0	0.3	2.9	0.1
Drama	8500	0.0	0.1	0.1	0.2	0.5	2.8	16.0	0.4
Economics	26105	0.0	0.0	0.0	0.1	0.0	0.1	5.5	0.1
Electronics	530	0.0	0.0	0.9	1.0	1.5	0.0	0.0	0.6
English lang	12490	0.0	0.0	0.0	0.0	0.2	2.3	34.3	0.3
English lang & lit	6740	0.0	0.0	0.0	0.1	0.2	2.0	25.0	0.3
English lit	34090	0.0	0.0	0.0	0.1	0.3	1.3	23.9	0.3
Environmental studies	740	0.0	0.0	0.0	0.0	0.0	0.9	2.4	0.3
Film studies	5100	0.0	0.3	0.3	0.1	0.7	2.3	38.5	0.7
French	6900	0.0	0.0	0.1	0.4	0.1	0.7	4.0	0.2
Further mathematics	11645	0.0	0.0	0.1	0.2	0.2	1.9	8.9	0.3
Geography	29545	0.0	0.0	0.0	0.0	0.1	0.2	8.1	0.2
Geology	1050	0.5	0.5	0.0	0.0	0.0	1.0	5.9	0.3
German	2485	0.0	0.0	0.5	0.4	0.3	0.9	0.0	0.3
History	43945	0.0	0.0	0.0	0.1	0.4	1.5	17.8	0.3

135 Based on data submitted to Ofqual by exam boards around a week before results were issued (any missing data is likely to be missing at random).
136 Subjects with fewer than 100 entries included in the overall total but not shown individually in each table. Entries rounded to nearest 5.

Italian	475	0.0	0.0	1.3	0.0	0.0	25.0	100.0	1.1
Latin	1015	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Law	10060	0.0	0.0	0.0	0.5	0.8	1.6	9.5	0.8
Mathematics	70860	0.0	0.0	0.1	0.1	0.2	0.5	2.7	0.2
Media studies	12740	0.0	0.1	0.1	0.4	0.5	3.6	25.3	0.6
Music	4710	0.0	0.0	0.0	0.2	0.1	0.3	8.9	0.3
Philosophy	2450	0.0	0.0	0.0	0.0	0.5	0.4	8.0	0.4
Physical education	9365	0.0	0.0	0.0	0.1	0.2	0.2	1.0	0.1
Physics	31980	0.0	0.0	0.0	0.1	0.1	0.0	0.9	0.1
Polish	330	0.0	0.0	0.0	0.0	4.0	16.7	40.0	1.5
Political studies	16215	0.0	0.0	0.0	0.0	0.1	0.7	8.4	0.2
Portuguese	250	0.0	0.0	0.0	0.0	0.0	9.1	26.7	2.0
Psychology	55915	0.0	0.0	0.0	0.1	0.1	0.5	4.4	0.2
Religious studies	14860	0.0	0.0	0.1	0.2	0.3	1.0	6.6	0.3
Russian	350	0.0	0.0	0.0	6.7	0.0	0.0	0.0	0.3
Sociology	31640	0.0	0.0	0.0	0.0	0.1	0.4	6.1	0.2
Spanish	6930	0.0	0.1	0.0	0.1	0.5	1.5	4.8	0.2
Statistics	555	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.2
Turkish	285	0.0	0.0	0.0	0.0	0.0	0.0	33.3	0.3
Urdu	150	0.0	1.9	3.0	0.0	22.2	0.0	50.0	3.3
Total	653305	0.0	0.0	0.0	0.1	0.2	0.6	4.8	0.2

Partial Absentees – AS

Percentage of 17-year-olds in England in 2019 that were partially absent by subject and grade

Subject	Entry	Subject level grade						
		A	B	C	D	E	U	Total
Accounting	240	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ancient history	130	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Arabic	130	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art	500	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: 3D studies	185	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: Fine art	1135	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: Graphics	465	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: Photography	1330	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art: Textiles	175	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Biology	7465	0.0	0.0	0.0	0.0	0.0	0.7	0.1
Business studies	4130	0.0	0.0	0.0	0.0	0.0	3.2	0.4
Chemistry	6615	0.0	0.0	0.0	0.0	0.0	0.7	0.1
Classical civilisation	280	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Computing	1690	0.0	0.0	0.0	0.0	0.0	1.3	0.2
Design and technology	635	0.0	0.0	0.0	0.0	0.0	2.9	0.3
Drama	590	0.0	0.0	0.6	0.0	5.9	0.0	0.3
Economics	3450	0.0	0.0	0.0	0.0	0.0	1.1	0.1
Electronics	175	0.0	0.0	0.0	0.0	0.0	33.3	0.6

English lang	1985	0.0	0.0	0.0	0.2	0.6	10.4	0.4
English lang & lit	765	0.0	0.0	0.0	0.0	0.0	17.2	0.7
English lit	3990	0.0	0.0	0.0	0.0	0.3	8.7	0.3
Environmental studies	225	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Film studies	640	0.0	0.0	0.0	0.0	2.2	23.5	0.8
French	900	0.0	0.0	0.0	0.8	0.0	6.5	0.6
Further mathematics	3535	0.0	0.0	0.0	0.0	0.5	1.4	0.1
Geography	2665	0.0	0.0	0.0	0.0	0.0	1.0	0.1
Geology	160	0.0	0.0	0.0	0.0	0.0	0.0	0.0
German	400	0.0	0.0	0.0	1.2	2.3	0.0	0.5
History	4265	0.0	0.0	0.0	0.0	0.0	2.1	0.1
Italian	105	0.0	0.0	12.5	0.0	33.3	0.0	1.9
Latin	105	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Law	1720	0.0	0.0	0.0	0.0	0.3	1.4	0.3
Mathematics	11430	0.0	0.0	0.0	0.1	0.1	1.0	0.2
Media studies	1445	0.0	0.3	0.2	0.3	0.6	17.3	1.2
Music	415	0.0	0.0	0.0	0.0	2.9	15.4	0.7
Philosophy	380	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Physical education	855	0.0	0.7	0.0	0.0	0.0	2.1	0.4
Physics	4720	0.0	0.0	0.0	0.0	0.0	0.6	0.1
Polish	190	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Political studies	1720	0.0	0.0	0.0	0.0	0.0	2.3	0.3
Psychology	8755	0.0	0.0	0.0	0.0	0.0	1.6	0.3
Religious studies	1950	0.0	0.2	0.2	0.0	0.6	4.4	0.5
Sociology	5615	0.0	0.0	0.0	0.0	0.0	3.7	0.5
Spanish	940	0.0	0.0	0.0	0.0	4.5	1.4	0.5
Turkish	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	89745	0.0	0.0	0.0	0.0	0.2	1.8	0.3

Partial Absentees – GCSE

Percentage of 16-year-olds in England in 2019 that were partially absent by subject and grade (post-reform 9 to 1 qualifications)

Subject	Entry	Subject level grade 9	8	7	6	5	4	3	2	1	U	Total
Ancient history	815	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6	0.5
Arabic	2205	0.0	0.0	0.6	1.1	0.8	2.9	1.2	5.7	0.0	16.0	2.2
Art	57785	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.7	2.7	8.6	0.2
Art: 3D studies	3800	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	2.8	3.4	0.2
Art: Fine art	64420	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.5	2.3	8.6	0.1
Art: Graphics	8535	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.4	2.2	4.8	0.2
Art: Photography	29955	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.9	3.3	9.9	0.2
Art: Textiles	9320	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.9	3.4	0.1
Astronomy	570	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.5	0.4
Bengali	440	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0	50.0	0.0	0.5
Biology	158885	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.7	7.9	16.7	0.2
Business studies	86390	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	2.2	14.6	0.3
Chemistry	155980	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	5.9	6.4	0.1
Chinese	2045	0.2	0.0	1.0	0.0	0.0	1.8	3.4	2.9	3.3	20.0	0.7
Citizenship studies	15855	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.8	7.4	24.3	1.3
Classical civilisation	3330	0.0	0.0	0.0	0.0	0.0	0.0	2.3	5.4	0.0	23.1	0.6
Classical Greek	1025	0.0	0.0	0.0	0.0	0.0	0.0	4.2	0.0	0.0	0.0	0.1
Combined science	383535	0.0	0.0	0.0	0.0	0.0	0.2	0.5	2.4	12.2	20.7	1.7
Computing	75190	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.3	4.2	0.2
Dance	8795	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.8	8.4	34.8	0.5
Design & technology	87640	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	3.7	27.4	0.7
Drama	56185	0.0	0.0	0.1	0.0	0.1	0.1	0.7	2.2	12.3	53.5	0.9
Economics	5815	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	10.3	0.1
Electronics	900	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.9	37.5	0.6
Engineering	2840	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	12.5	0.4

English language	546350	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	6.1	17.6	0.4
English literature	514145	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	2.0	12.8	0.3
Film studies	4475	0.8	0.0	0.3	0.1	0.0	0.2	1.3	3.4	10.8	55.8	2.8
Food prep & nutrition	43595	0.1	0.0	0.2	0.2	0.2	0.2	0.2	1.3	8.8	47.1	0.9
French	118440	0.0	0.0	0.0	0.0	0.1	0.1	0.3	1.4	6.5	6.9	0.4
Geography	246890	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	3.3	15.0	0.5
Geology	315	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
German	39600	0.0	0.0	0.2	0.2	0.1	0.2	0.4	2.0	7.4	12.4	0.6
Greek	230	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.3	0.0	26.5	4.3
Gujarati	195	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
History	256480	0.0	0.0	0.0	0.0	0.1	0.1	0.2	0.6	2.6	12.9	0.7
Italian	3425	0.0	0.0	0.9	1.2	0.9	1.5	0.5	1.4	30.8	25.8	0.8
Japanese	440	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.4	50.0	0.7
Latin	8705	0.0	0.0	0.0	0.2	0.0	0.3	0.0	0.0	3.2	8.3	0.1
Mathematics	547390	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.8	3.7	16.8	0.6
Media studies	30300	0.0	0.1	0.1	0.2	0.2	0.2	0.4	2.9	9.9	40.0	1.4
Modern Hebrew	105	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	1.0
Music	32915	0.0	0.0	0.0	0.0	0.0	0.1	0.4	1.4	5.0	28.1	0.5
Panjabi	480	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	100.0	26.1	1.7
Persian	180	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Physical education	77735	0.0	0.0	0.0	0.0	0.0	0.1	0.1	1.1	10.0	36.4	0.4
Physics	155190	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.6	5.6	7.0	0.1
Polish	3005	0.0	0.0	0.0	3.2	5.2	23.1	20.4	71.4	100.0	34.3	2.6
Portuguese	1355	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.3
Psychology	13965	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	1.3	14.4	0.5
Religious studies	220760	0.0	0.0	0.0	0.0	0.1	0.2	0.2	0.9	3.3	11.3	0.5
Russian	1325	0.0	1.4	1.8	1.8	2.9	8.0	0.0	57.1	0.0	80.0	1.6
Sociology	20215	0.0	0.1	0.0	0.0	0.0	0.0	0.2	0.5	4.1	14.3	0.5
Spanish	92110	0.0	0.0	0.1	0.1	0.1	0.2	0.3	1.6	10.4	8.4	0.6
Statistics	13135	0.0	0.0	0.0	0.0	0.0	0.3	0.3	0.9	14.0	14.4	1.0
Turkish	1005	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Urdu	2980	0.0	0.0	0.0	0.0	0.7	0.8	1.1	2.3	20.0	18.2	1.2
Total	4219775	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.8	4.3	15.4	0.5

Annex D: Prior-attainment match rates

Table D.1. GCSE – All Centres

All Centres		**Prior Attainment Match Rates (Percentage of centres)**										
Subject Grouping	**Number of centres**	**>0.9 to 1.0**	**>0.8 to 0.9**	**>0.7 to 0.8**	**>0.6 to 0.7**	**>0.5 to 0.6**	**>0.4 to 0.5**	**>0.3 to 0.4**	**>0.2 to 0.3**	**>0.1 to 0.2**	**>0.0 to 0.1**	**0**
Ancient History	57	35.1	17.5	10.5	3.5	3.5	5.3	1.8	5.3	3.5	1.8	12.3
Arabic	621	15.5	2.4	2.6	3.2	3.1	6.0	3.7	2.7	4.7	1.4	54.8
Art & Design: 3D Studies	340	36.2	21.8	8.8	6.5	2.4	5.3	4.4	3.2	2.9	1.2	7.4
Art & Design: Art, Craft and Design	2207	34.5	22.4	10.0	4.7	2.9	4.3	4.0	2.5	3.2	1.6	10.1
Art & Design: Critical and Contextual Studies	6	33.3	33.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.7	16.7
Art & Design: Fine Art	2443	36.8	21.3	9.4	4.1	2.7	3.7	3.8	2.9	3.6	2.5	9.2
Art & Design: Graphics	446	46.0	23.5	6.3	2.7	1.3	2.5	4.0	0.7	1.6	0.4	11.0
Art & Design: Photography	1392	43.5	21.3	8.9	4.0	2.1	2.7	2.7	2.2	1.7	0.8	10.0
Art & Design: Textiles	774	44.1	22.0	10.5	3.7	2.6	2.3	2.8	2.2	1.7	0.6	7.5
Astronomy	164	20.7	7.3	2.4	3.7	0.6	4.3	1.8	2.4	4.3	2.4	50.0
Bengali	83	16.9	1.2	6.0	3.6	2.4	4.8	3.6	1.2	0.0	0.0	60.2
Biblical Hebrew	23	8.7	0.0	0.0	4.3	0.0	0.0	0.0	0.0	0.0	0.0	87.0
Biology	3931	50.3	20.5	6.0	2.8	1.8	3.2	2.8	2.2	2.1	1.1	7.2
Business	2016	50.4	23.5	6.9	3.2	1.8	3.4	2.2	1.7	1.0	1.3	4.5
Chemistry	3557	54.9	21.6	5.5	2.2	1.6	2.5	2.6	1.9	1.5	1.2	4.3
Chinese	495	18.4	5.7	2.8	1.6	0.2	3.8	2.6	0.6	3.2	2.0	59.0
Citizenship Studies	594	36.2	17.5	10.3	4.9	3.0	4.2	3.7	2.7	2.2	1.9	13.5
Classical Civilisation	259	19.7	6.6	2.7	4.6	3.9	7.3	10.8	10.4	10.8	6.6	16.6
Classical Greek	200	8.5	4.0	1.5	3.0	1.5	9.5	6.5	10.0	7.5	3.0	45.0
Combined Science	4075	42.3	27.6	8.8	3.9	2.2	2.9	2.6	2.4	1.8	1.2	4.2
Computing	3037	45.3	25.7	9.2	3.0	1.9	2.7	2.0	1.8	1.7	1.1	5.6
Dance	689	56.5	17.1	6.1	3.3	1.6	1.7	2.3	1.6	1.9	0.3	7.5
Design and Technology	2763	46.8	23.5	7.9	2.8	2.0	3.3	2.7	2.3	1.7	1.7	5.3
Drama	2601	45.0	23.1	7.6	3.2	1.7	3.1	3.2	3.0	3.3	1.8	4.8
Economics	237	38.8	27.8	8.0	3.8	0.8	2.5	1.3	3.0	1.3	2.1	10.5
Electronics	58	27.6	13.8	6.9	3.4	5.2	3.4	3.4	3.4	3.4	3.4	25.9
Engineering	97	61.9	18.6	8.2	1.0	0.0	0.0	2.1	2.1	0.0	0.0	6.2
English Language	5292	29.0	25.8	9.2	4.4	2.9	3.5	3.3	2.4	2.0	3.3	14.1
English Literature	4236	42.7	26.2	7.6	3.9	2.4	4.2	3.1	2.0	1.3	2.5	4.2
Film Studies	262	42.7	25.2	4.6	4.2	1.9	3.1	2.3	1.1	1.1	0.0	13.7
Food Prep and Nutrition	2072	48.5	24.0	9.3	3.2	1.8	2.7	1.9	2.2	1.6	0.9	4.0
French	3362	46.7	22.7	8.7	3.4	1.7	2.7	2.3	2.1	2.3	1.4	5.9
Geography	3877	47.9	24.8	7.9	3.2	1.8	2.9	3.1	2.3	1.9	1.4	2.6
Geology	35	34.3	8.6	2.9	2.9	0.0	2.9	0.0	2.9	0.0	5.7	40.0
German	1511	54.1	15.6	4.4	2.2	1.0	3.1	2.3	1.6	1.5	0.7	13.7
Greek	164	9.1	0.0	0.6	0.0	0.0	1.2	0.6	0.0	0.0	1.8	86.6
Gujarati	45	13.3	0.0	0.0	2.2	0.0	4.4	6.7	0.0	2.2	4.4	66.7
History	3910	54.0	23.2	6.0	2.2	1.6	3.1	2.3	1.7	1.3	0.8	3.8
Italian	736	17.0	1.9	0.7	1.6	0.4	5.3	3.4	1.6	2.0	1.2	64.8
Japanese	134	17.2	1.5	1.5	1.5	0.0	3.7	2.2	0.0	2.2	1.5	68.7
Latin	608	18.8	7.1	6.4	3.3	4.3	6.9	10.5	7.7	9.2	6.4	19.4
Mathematics	5367	26.3	26.4	10.5	5.0	3.1	3.6	2.7	2.4	1.9	3.5	14.8
Media Studies	1017	51.8	23.5	5.8	2.9	0.9	2.2	1.2	0.9	1.2	1.8	7.9
Modern Hebrew	28	7.1	0.0	7.1	3.6	0.0	0.0	3.6	10.7	3.6	3.6	60.7
Music	2738	43.3	19.0	9.0	3.9	1.7	4.1	2.6	2.9	3.0	1.2	9.5
Panjabi	105	27.6	4.8	2.9	1.9	1.0	5.7	1.9	1.9	1.9	2.9	47.6
Persian	125	23.2	0.0	0.0	1.6	0.0	4.8	0.8	0.8	0.8	0.0	68.0
Physical Education	2671	52.3	16.8	6.4	2.2	1.8	3.4	3.1	2.8	2.8	1.8	6.6
Physical Education (SC)	119	21.0	5.0	3.4	6.7	4.2	6.7	7.6	0.8	3.4	1.7	39.5
Physics	3539	55.3	21.6	5.4	2.4	1.6	2.5	2.7	1.7	1.6	1.1	4.1
Polish	836	32.5	1.9	4.1	3.7	2.2	11.2	4.8	1.9	1.6	0.2	35.9
Portuguese	528	17.0	0.6	3.0	5.3	0.8	9.5	5.7	3.4	2.1	0.8	51.9
Psychology	659	40.1	14.0	3.6	1.5	0.9	1.7	2.3	1.7	3.0	2.9	28.4
Religious Studies	2995	43.8	22.3	7.1	4.2	2.7	3.4	3.0	2.6	2.2	1.9	6.7
Religious Studies (SC)	348	28.4	10.3	7.8	4.0	1.7	4.9	4.0	4.0	3.2	3.2	28.4
Russian	492	21.3	0.4	1.0	1.0	0.4	5.1	2.6	1.0	1.4	0.8	64.8
Sociology	686	46.6	23.5	7.4	1.6	1.2	1.2	1.9	0.4	1.3	2.9	12.0
Spanish	2850	43.2	20.2	9.1	3.8	2.4	3.8	3.5	2.1	1.8	1.4	8.7
Statistics	780	36.9	12.6	5.9	3.3	1.5	2.9	1.7	1.2	1.3	3.6	29.1

Turkish	301	38.2	4.7	3.0	1.7	1.7	7.0	1.7	1.0	0.7	0.0	40.5
Urdu	309	17.8	8.7	8.4	6.8	2.9	5.5	2.3	2.6	1.3	1.3	42.4

Table D.2. GCSE – At least 15 Entries

Centres with at least 15 entries in 2020		**Prior Attainment Match Rates (Percentage of centres)**										
Subject Grouping	**Number of centres**	**>0.9 to 1.0**	**>0.8 to 0.9**	**>0.7 to 0.8**	**>0.6 to 0.7**	**>0.5 to 0.6**	**>0.4 to 0.5**	**>0.3 to 0.4**	**>0.2 to 0.3**	**>0.1 to 0.2**	**>0.0 to 0.1**	**0**
Ancient History	31	41.9	22.6	12.9	3.2	3.2	0.0	3.2	9.7	0.0	3.2	0.0
Arabic	71	7.0	12.7	5.6	5.6	12.7	2.8	1.4	4.2	7.0	8.5	32.4
Art & Design: 3D Studies	182	40.1	32.4	8.2	3.8	1.6	3.8	2.2	1.1	1.6	1.1	3.8
Art & Design: Art, Craft and Design	1486	40.2	29.8	10.5	4.4	2.0	2.5	2.2	1.5	2.1	1.5	3.2
Art & Design: Critical and Contextual Studies	4	25.0	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	25.0	25.0
Art & Design: Fine Art	1801	42.0	26.9	9.2	3.9	2.3	2.2	2.8	2.3	2.5	2.4	3.4
Art & Design: Graphics	262	50.8	31.7	5.7	2.7	1.1	1.1	2.7	0.0	0.4	0.4	3.4
Art & Design: Photography	967	50.3	26.7	9.7	4.0	1.8	0.9	1.3	1.9	0.8	0.5	2.1
Art & Design: Textiles	339	48.4	28.0	9.7	2.7	1.2	1.2	2.7	1.2	1.8	0.3	2.9
Astronomy	26	15.4	19.2	3.8	0.0	3.8	3.8	0.0	3.8	3.8	11.5	34.6
Bengali	8	0.0	12.5	50.0	12.5	12.5	0.0	0.0	0.0	0.0	0.0	12.5
Biblical Hebrew	16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Biology	3164	55.5	23.8	5.3	2.2	1.5	2.1	2.1	1.5	1.7	1.2	3.1
Business	1740	52.9	26.4	6.8	3.0	1.6	2.4	1.7	1.3	0.9	1.3	1.7
Chemistry	3024	58.6	23.9	5.1	1.7	1.4	1.8	2.2	1.3	1.3	1.1	1.5
Chinese	77	14.3	15.6	3.9	1.3	1.3	2.6	5.2	1.3	9.1	10.4	35.1
Citizenship Studies	354	38.4	23.4	10.7	3.7	2.5	2.8	2.3	1.4	1.7	2.5	10.5
Classical Civilisation	107	24.3	10.3	2.8	5.6	2.8	4.7	14.0	11.2	9.3	8.4	6.5
Classical Greek	12	8.3	8.3	0.0	0.0	0.0	8.3	8.3	16.7	8.3	25.0	16.7
Combined Science	3461	45.7	31.2	9.0	3.4	1.7	1.6	1.8	1.8	1.3	1.2	1.2
Computing	2250	49.5	29.9	9.4	2.7	1.5	1.2	1.2	1.0	0.6	0.8	2.2
Dance	244	64.8	20.5	3.3	3.3	0.4	0.4	0.4	0.8	1.2	0.4	4.5
Design and Technology	2245	49.9	26.2	7.5	2.3	1.6	2.5	2.2	1.9	1.4	1.8	2.7
Drama	1821	51.0	26.1	6.6	2.5	1.2	2.0	2.1	2.1	2.6	1.9	1.9
Economics	176	44.3	32.4	9.1	2.8	0.0	2.3	1.1	1.7	0.0	2.3	4.0
Electronics	33	33.3	15.2	6.1	3.0	6.1	3.0	6.1	3.0	6.1	6.1	12.1
Engineering	83	68.7	21.7	6.0	1.2	0.0	0.0	1.2	0.0	0.0	0.0	1.2
English Language	4386	32.3	30.0	9.4	3.9	2.6	2.2	2.6	2.1	1.3	3.9	9.6
English Literature	3633	45.3	29.5	7.5	3.2	2.1	3.0	2.6	1.7	1.0	2.9	1.3
Film Studies	161	46.0	29.2	5.0	2.5	1.9	1.9	0.6	1.2	1.2	0.0	10.6
Food Prep and Nutrition	1443	53.6	27.1	8.5	2.7	1.2	1.4	1.1	1.2	0.4	0.7	1.9
French	2535	50.7	27.3	9.2	2.4	1.1	1.5	1.6	1.4	1.8	1.5	1.5
Geography	3505	50.6	26.6	8.1	2.8	1.4	2.2	2.5	2.1	1.7	1.3	0.7
Geology	15	40.0	13.3	0.0	6.7	0.0	6.7	0.0	6.7	0.0	6.7	20.0
German	930	67.3	20.3	4.5	1.3	0.4	1.7	0.9	1.1	0.8	0.9	0.9
Greek	9	0.0	0.0	11.1	0.0	0.0	0.0	0.0	0.0	0.0	33.3	55.6
Gujarati	8	0.0	0.0	0.0	12.5	0.0	0.0	0.0	0.0	12.5	25.0	50.0
History	3395	58.3	25.9	5.9	1.7	1.0	1.8	1.4	1.2	0.9	0.9	0.9
Italian	59	42.4	22.0	6.8	3.4	3.4	0.0	1.7	0.0	3.4	8.5	8.5
Japanese	15	33.3	13.3	13.3	6.7	0.0	6.7	0.0	0.0	6.7	6.7	13.3
Latin	237	16.5	11.4	9.3	1.7	6.8	5.9	10.5	10.1	8.9	13.1	5.9
Mathematics	4407	29.2	31.3	11.1	4.8	2.6	2.1	1.7	1.7	1.1	4.0	10.4
Media Studies	825	54.4	25.9	5.8	2.9	0.2	1.3	1.0	0.7	1.1	1.7	4.8
Modern Hebrew	16	6.3	0.0	12.5	6.3	0.0	0.0	6.3	12.5	6.3	6.3	43.8
Music	1003	49.4	27.3	9.1	3.3	1.0	1.2	1.3	1.5	2.2	1.7	2.1
Panjabi	11	9.1	45.5	0.0	0.0	9.1	0.0	0.0	9.1	9.1	9.1	9.1
Persian	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Physical Education	2052	60.5	19.3	5.6	2.1	1.0	1.9	1.8	1.8	1.8	1.7	2.5
Physical Education (SC)	13	7.7	15.4	7.7	7.7	0.0	0.0	0.0	0.0	0.0	15.4	46.2
Physics	3011	58.8	24.0	4.9	1.7	1.5	1.9	2.1	1.2	1.3	1.2	1.4
Polish	21	4.8	23.8	14.3	14.3	9.5	4.8	0.0	4.8	0.0	4.8	19.0
Portuguese	15	0.0	0.0	6.7	6.7	0.0	6.7	26.7	20.0	6.7	13.3	13.3
Psychology	439	49.0	19.8	4.1	1.6	0.7	0.7	1.1	1.1	3.6	3.2	15.0
Religious Studies	2368	45.6	25.2	6.9	3.9	2.6	3.0	2.3	2.2	1.8	2.0	4.6
Religious Studies (SC)	191	28.8	13.1	7.9	3.7	1.0	2.6	2.6	3.7	2.6	5.2	28.8
Russian	23	21.7	8.7	4.3	0.0	0.0	4.3	17.4	4.3	4.3	8.7	26.1
Sociology	566	48.9	26.7	7.8	1.6	0.9	0.5	1.6	0.4	1.1	3.2	7.4
Spanish	2135	48.4	24.7	9.8	3.7	2.2	2.5	2.4	1.6	1.3	1.5	1.9
Statistics	418	37.1	17.2	5.3	1.9	1.2	2.4	1.2	1.7	1.2	6.2	24.6
Turkish	23	13.0	34.8	21.7	4.3	0.0	8.7	4.3	0.0	4.3	0.0	8.7
Urdu	79	15.2	21.5	15.2	12.7	7.6	3.8	1.3	1.3	1.3	3.8	16.5

Table D.3. AS – All Centres

All Centres		**Prior Attainment Match Rates (Percentage of centres)**										
Subject Grouping	**Number of centres**	**>0.9 to 1.0**	**>0.8 to 0.9**	**>0.7 to 0.8**	**>0.6 to 0.7**	**>0.5 to 0.6**	**>0.4 to 0.5**	**>0.3 to 0.4**	**>0.2 to 0.3**	**>0.1 to 0.2**	**>0.0 to 0.1**	**0**
Accounting	37	10.8	13.5	18.9	8.1	2.7	5.4	0.0	0.0	0.0	0.0	40.5
Ancient History	10	40.0	0.0	10.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	40.0
Art & Design: 3D Studies	20	25.0	5.0	15.0	5.0	0.0	0.0	5.0	0.0	0.0	0.0	45.0
Art & Design: Art, Craft and Design	105	27.6	3.8	9.5	1.9	1.0	2.9	1.0	0.0	0.0	0.0	52.4
Art & Design: Fine Art	229	30.6	4.8	3.1	1.3	0.4	3.5	1.7	0.4	0.9	0.0	53.3
Art & Design: Graphics	49	26.5	6.1	10.2	0.0	2.0	6.1	0.0	0.0	6.1	0.0	42.9
Art & Design: Photography	215	20.5	8.8	6.0	3.3	0.0	5.6	3.3	0.9	1.9	0.0	49.8
Art & Design: Textiles	48	31.3	0.0	6.3	2.1	2.1	6.3	0.0	2.1	2.1	0.0	47.9
Biology	628	34.7	7.2	2.7	3.0	0.8	4.1	2.5	0.8	0.8	0.0	43.3
Business Studies	383	28.2	8.4	4.4	2.1	2.1	4.2	1.6	1.3	1.3	0.3	46.2
Chemistry	601	38.1	5.0	4.3	2.7	1.5	4.0	1.7	1.3	0.8	0.3	40.3
Chinese	83	8.4	0.0	0.0	0.0	0.0	0.0	0.0	1.2	1.2	0.0	89.2
Classical Civilisation	38	23.7	21.1	5.3	2.6	5.3	15.8	2.6	0.0	0.0	0.0	23.7
Computing	248	40.7	5.6	6.0	2.8	0.4	4.0	0.8	0.4	0.4	0.4	38.3
D&T: Product Design	82	36.6	6.1	7.3	2.4	2.4	2.4	1.2	0.0	0.0	0.0	41.5
Drama & Theatre Studies	80	56.3	6.3	2.5	0.0	1.3	2.5	0.0	0.0	1.3	1.3	28.7
Economics	313	32.6	6.4	3.2	2.2	2.2	2.6	3.5	0.6	1.0	1.0	44.7
Electronics	26	46.2	11.5	3.8	3.8	0.0	3.8	0.0	0.0	0.0	0.0	30.8
English Language	165	34.5	7.3	7.9	3.6	1.2	4.2	4.2	2.4	0.0	0.6	33.9
English Language & Literature	80	30.0	11.3	8.8	2.5	2.5	1.3	2.5	2.5	1.3	1.3	36.3
English Literature	272	43.8	8.1	3.3	2.2	1.8	3.7	0.0	1.1	0.7	1.1	34.2
Environmental Studies	24	20.8	0.0	16.7	4.2	0.0	0.0	0.0	0.0	4.2	0.0	54.2
Film Studies	74	20.3	12.2	9.5	9.5	2.7	4.1	2.7	9.5	0.0	0.0	29.7
French	238	46.6	1.3	3.8	1.7	1.3	3.4	0.0	0.4	0.8	0.0	40.8
Further Mathematics	816	35.9	5.8	2.7	2.9	1.1	2.8	1.1	0.6	0.2	0.0	46.8
Geography	243	47.7	6.2	4.5	3.3	0.8	4.9	1.2	0.0	0.0	0.0	31.3
Geology	25	36.0	8.0	16.0	4.0	0.0	4.0	0.0	0.0	0.0	0.0	32.0
German	130	51.5	0.8	1.5	0.8	0.0	2.3	0.8	1.5	0.8	0.0	40.0
History	351	41.0	6.3	4.6	3.4	2.3	4.6	1.7	1.1	0.3	0.3	34.5
Italian	33	39.4	0.0	3.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	54.5
Law	102	26.5	7.8	4.9	6.9	4.9	2.9	2.9	0.0	0.0	0.0	43.1
Mathematics	1031	22.5	4.6	2.8	1.8	0.9	4.6	2.5	1.8	1.6	0.6	56.4
Media Studies	103	34.0	8.7	3.9	6.8	2.9	4.9	4.9	2.9	1.0	1.0	29.1
Music	84	48.8	1.2	1.2	4.8	0.0	1.2	1.2	0.0	0.0	0.0	41.7
Music Technology	34	32.4	2.9	5.9	2.9	2.9	0.0	0.0	0.0	0.0	0.0	52.9
Philosophy	48	29.2	6.3	2.1	6.3	2.1	4.2	4.2	0.0	0.0	2.1	43.8
Physical Education	125	42.4	4.8	2.4	0.8	0.0	3.2	0.0	1.6	0.0	0.0	44.8
Physics	570	39.8	5.6	2.6	2.6	0.4	4.6	1.8	1.1	0.5	0.2	40.9
Politics	205	36.6	7.8	5.9	3.4	1.0	3.4	2.0	2.4	1.0	0.0	36.6
Psychology	649	29.0	6.9	4.6	2.8	2.6	5.5	2.2	1.1	0.8	0.3	44.2
Religious Studies	202	41.1	5.0	4.0	3.0	1.0	6.4	1.0	1.5	0.0	0.5	36.6
Sociology	443	25.5	7.0	7.7	3.6	3.2	5.0	1.6	0.5	1.1	0.9	44.0
Spanish	221	43.0	3.2	3.2	3.6	0.9	3.6	1.4	0.0	0.0	0.9	40.3
Statistics	28	14.3	7.1	0.0	0.0	10.7	0.0	0.0	0.0	0.0	0.0	67.9

Table D.4. AS – At least 15 Entries

Centres with at least 15 entries in 2020		**Prior Attainment Match Rates (Percentage of centres)**										
Subject Grouping	**Number of centres**	**>0.9 to 1.0**	**>0.8 to 0.9**	**>0.7 to 0.8**	**>0.6 to 0.7**	**>0.5 to 0.6**	**>0.4 to 0.5**	**>0.3 to 0.4**	**>0.2 to 0.3**	**>0.1 to 0.2**	**>0.0 to 0.1**	**0**
Accounting	8	12.5	37.5	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ancient History	5	60.0	0.0	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
Art & Design: 3D Studies	3	33.3	33.3	33.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art & Design: Art, Craft and Design	2	50.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art & Design: Fine Art	14	35.7	21.4	21.4	14.3	0.0	0.0	0.0	0.0	0.0	0.0	7.1
Art & Design: Graphics	8	12.5	25.0	37.5	0.0	12.5	0.0	0.0	0.0	0.0	0.0	12.5
Art & Design: Photography	15	26.7	40.0	20.0	6.7	0.0	0.0	0.0	0.0	0.0	0.0	6.7
Art & Design: Textiles	0	-	-	-	-	-	-	-	-	-	-	-
Biology	121	50.4	20.7	6.6	4.1	3.3	0.8	5.0	1.7	0.8	0.0	6.6
Business Studies	73	37.0	30.1	9.6	6.8	5.5	2.7	0.0	0.0	0.0	1.4	6.8
Chemistry	93	52.7	14.0	9.7	2.2	4.3	1.1	5.4	3.2	1.1	1.1	5.4
Chinese	12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Classical Civilisation	4	25.0	25.0	0.0	0.0	25.0	0.0	0.0	0.0	0.0	0.0	25.0
Computing	25	48.0	20.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	8.0
D&T: Product Design	7	57.1	28.6	14.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Drama & Theatre Studies	9	44.4	33.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.1	11.1
Economics	57	43.9	21.1	5.3	0.0	3.5	1.8	5.3	1.8	1.8	3.5	12.3
Electronics	5	40.0	40.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
English Language	24	20.8	33.3	16.7	8.3	0.0	12.5	0.0	0.0	0.0	4.2	4.2
English Language & Literature	19	26.3	31.6	10.5	5.3	10.5	0.0	5.3	0.0	5.3	5.3	0.0
English Literature	60	58.3	18.3	8.3	1.7	3.3	1.7	0.0	0.0	0.0	5.0	3.3
Environmental Studies	5	20.0	0.0	60.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
Film Studies	21	19.0	23.8	9.5	19.0	0.0	9.5	0.0	9.5	0.0	0.0	9.5
French	5	40.0	0.0	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
Further Mathematics	69	46.4	26.1	1.4	4.3	1.4	1.4	1.4	2.9	0.0	0.0	14.5
Geography	39	76.9	12.8	5.1	0.0	0.0	2.6	0.0	0.0	0.0	0.0	2.6
Geology	4	25.0	25.0	25.0	25.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
German	2	50.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0
History	64	57.8	17.2	6.3	4.7	6.3	1.6	0.0	0.0	0.0	1.6	4.7
Italian	0	-	-	-	-	-	-	-	-	-	-	-
Law	35	45.7	20.0	14.3	8.6	5.7	2.9	0.0	0.0	0.0	0.0	2.9
Mathematics	158	47.5	13.9	7.0	1.9	2.5	2.5	5.1	3.8	1.9	1.9	12.0
Media Studies	31	38.7	19.4	0.0	12.9	6.5	6.5	0.0	3.2	0.0	3.2	9.7
Music	0	-	-	-	-	-	-	-	-	-	-	-
Music Technology	2	0.0	0.0	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0
Philosophy	10	30.0	20.0	10.0	0.0	10.0	0.0	10.0	0.0	0.0	10.0	10.0
Physical Education	14	71.4	14.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.3
Physics	66	47.0	19.7	3.0	4.5	1.5	1.5	3.0	6.1	3.0	1.5	9.1
Politics	26	34.6	34.6	7.7	11.5	3.8	3.8	0.0	0.0	0.0	0.0	3.8
Psychology	138	47.1	19.6	12.3	5.8	5.8	2.9	0.7	0.0	0.7	0.7	4.3
Religious Studies	29	51.7	17.2	6.9	10.3	3.4	0.0	3.4	0.0	0.0	0.0	6.9
Sociology	103	35.0	19.4	19.4	5.8	8.7	3.9	1.9	0.0	0.0	2.9	2.9
Spanish	8	12.5	12.5	12.5	25.0	0.0	0.0	0.0	0.0	0.0	12.5	25.0
Statistics	4	25.0	25.0	0.0	0.0	25.0	0.0	0.0	0.0	0.0	0.0	25.0

Table D.5. A Level – All Centres

All Centres		**Prior Attainment Match Rates (Percentage of centres)**										
Subject Grouping	**Number of centres**	**>0.9 to 1.0**	**>0.8 to 0.9**	**>0.7 to 0.8**	**>0.6 to 0.7**	**>0.5 to 0.6**	**>0.4 to 0.5**	**>0.3 to 0.4**	**>0.2 to 0.3**	**>0.1 to 0.2**	**>0.0 to 0.1**	**0**
Accounting	160	50.6	16.9	5.0	3.1	3.1	3.1	3.1	1.3	0.6	0.0	13.1
Ancient History	65	44.6	13.8	9.2	6.2	1.5	1.5	1.5	0.0	0.0	3.1	18.5
Arabic	120	25.8	0.0	0.0	3.3	0.0	6.7	1.7	1.7	2.5	0.0	58.3
Art & Design: 3D Studies	147	62.6	12.9	5.4	4.8	0.0	3.4	1.4	0.7	0.7	0.0	8.2
Art & Design: Art, Craft and Design	801	61.7	10.0	7.4	3.7	1.2	2.7	1.4	1.0	0.6	0.2	10.0
Art & Design: Critical and Contextual Studies	20	40.0	5.0	5.0	5.0	5.0	15.0	0.0	0.0	10.0	0.0	15.0
Art & Design: Fine Art	1671	60.2	9.8	7.7	4.2	1.9	2.9	1.7	0.7	1.0	0.2	9.8
Art & Design: Graphics	397	59.9	14.1	9.8	2.8	1.5	2.8	1.0	0.3	0.0	0.0	7.8
Art & Design: Photography	1235	58.1	12.1	8.3	5.8	2.2	3.4	1.5	0.8	0.6	0.2	7.0
Art & Design: Textiles	483	69.6	6.8	4.8	4.8	1.2	4.3	1.2	0.4	0.2	0.4	6.2
Biology	2555	50.3	20.5	9.0	4.5	2.3	2.5	1.9	1.8	1.6	1.1	4.5
Business Studies	1740	54.3	16.8	8.2	4.9	2.5	2.8	1.6	1.3	0.9	0.7	6.0
Chemistry	2509	49.0	17.9	9.8	4.8	2.9	3.2	1.9	2.2	1.7	1.3	5.4
Chinese	423	5.7	0.0	0.5	0.0	0.0	3.5	3.3	0.9	6.6	4.7	74.7
Classical Civilisation	379	52.8	10.3	7.9	6.6	2.6	4.2	1.6	1.3	1.6	0.3	10.8
Classical Greek	88	61.4	0.0	4.5	3.4	0.0	5.7	4.5	1.1	1.1	0.0	18.2
Computing	1404	62.0	12.1	7.8	3.9	1.1	3.3	1.5	0.9	0.5	0.1	6.8
D&T: Fashion and Textiles	153	82.4	5.9	2.0	2.6	0.7	2.6	0.0	0.0	0.7	0.7	2.6
D&T: Product Design	1255	68.0	8.8	6.4	3.7	1.4	2.1	1.4	0.6	0.9	0.2	6.5
D&T: Design Engineering	47	66.0	10.6	4.3	2.1	0.0	2.1	0.0	0.0	2.1	0.0	12.8
Dance	214	78.0	3.7	6.5	3.3	0.9	1.9	0.5	0.0	0.0	0.0	5.1
Drama & Theatre Studies	1309	66.4	8.4	7.1	3.7	1.3	3.0	1.1	0.6	1.1	0.2	7.2
Economics	1657	47.7	16.5	8.3	4.5	3.1	2.7	2.7	2.5	2.3	1.7	8.1
Electronics	64	62.5	9.4	12.5	6.3	1.6	1.6	0.0	3.1	0.0	0.0	3.1
English Language	984	64.9	17.5	5.4	3.0	0.7	2.1	0.5	0.4	0.7	0.2	4.5
English Language & Literature	563	58.1	18.1	6.6	3.9	1.2	1.6	0.9	0.4	0.9	0.2	8.2
English Literature	2451	60.3	16.0	7.1	3.7	1.4	2.5	1.1	0.7	1.1	0.6	5.5
Environmental Studies	60	48.3	21.7	11.7	8.3	0.0	0.0	0.0	1.7	0.0	0.0	8.3
Film Studies	511	57.5	17.8	9.0	4.5	2.3	3.1	1.6	0.2	0.4	0.0	3.5
French	1519	61.4	6.2	7.4	3.1	1.4	3.8	2.2	1.4	1.1	0.7	11.5
Further Mathematics	1924	58.5	8.1	6.2	3.9	1.6	3.6	2.4	1.8	1.9	0.8	11.2
Geography	2100	68.3	12.9	5.7	3.0	1.2	2.2	0.8	0.6	0.8	0.5	3.9
Geology	105	49.5	22.9	12.4	1.9	2.9	2.9	1.0	1.0	1.0	0.0	4.8
German	791	61.3	4.6	4.4	4.2	0.4	5.1	1.6	0.8	0.5	0.0	17.2
History	2470	64.0	13.8	7.1	2.5	1.4	1.9	1.1	0.9	1.3	0.6	5.3
History of Art	85	38.8	3.5	7.1	4.7	1.2	9.4	3.5	5.9	3.5	0.0	22.4
Italian	293	43.3	1.4	1.0	3.1	0.3	6.1	1.7	0.7	1.4	0.3	40.6
Japanese	54	27.8	0.0	1.9	0.0	0.0	1.9	1.9	0.0	0.0	0.0	66.7
Latin	308	64.6	2.9	3.2	2.6	0.6	7.5	3.6	0.6	2.3	0.6	11.4
Law	536	59.0	19.0	7.1	4.5	2.8	2.4	0.9	0.2	0.4	0.0	3.7
Mathematics	2730	49.4	16.2	8.6	4.8	2.8	2.8	2.7	1.9	2.5	1.5	6.8
Media Studies	993	60.9	16.3	11.3	2.9	1.9	2.4	0.3	0.4	0.1	0.1	3.3
Music	1059	70.4	5.0	5.0	2.8	0.7	4.7	0.8	0.2	0.2	0.0	10.2
Music Technology	240	56.7	10.8	8.8	4.2	1.3	5.8	2.5	0.0	0.4	0.0	9.6
Panjabi	18	33.3	0.0	5.6	5.6	0.0	5.6	5.6	0.0	0.0	0.0	44.4
Philosophy	240	46.3	17.1	11.3	4.2	1.7	3.3	2.5	1.3	1.3	1.3	10.0
Physical Education	1290	70.2	10.6	5.2	2.6	0.5	1.6	0.6	0.2	0.6	0.0	7.9
Physics	2432	53.1	15.0	8.1	4.6	2.5	3.5	2.2	1.6	2.3	0.8	6.2
Polish	214	38.3	1.9	1.4	0.5	0.5	3.7	1.4	0.5	1.4	0.0	50.5
Politics	1341	50.5	17.1	9.8	4.1	2.1	2.2	1.8	1.0	2.0	1.1	8.4
Portuguese	78	37.2	0.0	0.0	2.6	0.0	10.3	2.6	1.3	0.0	1.3	44.9
Psychology	2403	55.8	19.0	7.7	4.3	2.3	2.7	1.2	0.9	1.2	0.5	4.5
Religious Studies	1487	58.9	16.8	9.3	3.1	2.5	2.6	1.4	0.7	0.6	0.0	4.0
Russian	253	25.7	0.8	0.4	0.0	0.0	4.0	4.7	0.4	1.6	1.2	61.3
Sociology	1715	56.7	20.6	9.6	4.0	1.9	2.0	0.5	0.4	0.3	0.3	3.7
Spanish	1434	55.2	9.0	7.9	4.9	1.4	4.5	1.5	1.5	1.4	0.6	12.1
Statistics	55	41.8	16.4	12.7	7.3	1.8	1.8	1.8	0.0	0.0	0.0	16.4
Turkish	85	40.0	4.7	1.2	2.4	0.0	2.4	0.0	1.2	0.0	0.0	48.2
Urdu	56	32.1	3.6	1.8	3.6	1.8	3.6	1.8	0.0	0.0	1.8	50.0

Table D.6. A Level – At least 15 Entries

Centres with at least 15 entries in 2020		**Prior Attainment Match Rates (Percentage of centres)**										
Subject Grouping	**Number of centres**	**>0.9 to 1.0**	**>0.8 to 0.9**	**>0.7 to 0.8**	**>0.6 to 0.7**	**>0.5 to 0.6**	**>0.4 to 0.5**	**>0.3 to 0.4**	**>0.2 to 0.3**	**>0.1 to 0.2**	**>0.0 to 0.1**	**0**
Accounting	47	55.3	29.8	8.5	0.0	2.1	2.1	0.0	0.0	0.0	0.0	2.1
Ancient History	16	31.3	37.5	25.0	0.0	6.3	0.0	0.0	0.0	0.0	0.0	0.0
Arabic	2	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0
Art & Design: 3D Studies	28	53.6	32.1	10.7	3.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Art & Design: Art, Craft and Design	57	56.1	21.1	8.8	3.5	0.0	0.0	0.0	1.8	1.8	0.0	7.0
Art & Design: Critical and Contextual Studies	3	0.0	33.3	0.0	33.3	0.0	33.3	0.0	0.0	0.0	0.0	0.0
Art & Design: Fine Art	181	61.3	21.5	7.7	2.8	1.1	0.0	0.6	0.6	2.2	0.6	1.7
Art & Design: Graphics	94	51.1	25.5	16.0	3.2	1.1	1.1	0.0	0.0	0.0	0.0	2.1
Art & Design: Photography	191	53.9	24.6	9.9	6.8	1.6	0.5	0.5	1.0	0.0	0.5	0.5
Art & Design: Textiles	36	58.3	19.4	8.3	2.8	0.0	2.8	0.0	0.0	0.0	5.6	2.8
Biology	1361	54.3	24.1	7.4	2.8	1.8	1.3	1.5	1.8	1.8	1.4	1.8
Business Studies	820	59.3	20.4	7.9	3.4	2.7	1.6	0.7	0.7	0.9	1.3	1.1
Chemistry	1123	53.3	21.5	8.5	2.7	1.9	2.1	1.5	2.7	2.0	1.8	2.2
Chinese	57	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	22.8	70.2
Classical Civilisation	41	51.2	29.3	12.2	2.4	2.4	0.0	0.0	2.4	0.0	0.0	0.0
Classical Greek	0	-	-	-	-	-	-	-	-	-	-	-
Computing	164	61.6	28.0	6.7	2.4	0.6	0.0	0.0	0.0	0.6	0.0	0.0
D&T: Fashion and Textiles	2	50.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0
D&T: Product Design	85	72.9	11.8	8.2	0.0	1.2	1.2	0.0	1.2	1.2	0.0	2.4
D&T: Design Engineering	3	66.7	33.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Dance	9	66.7	11.1	11.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.1
Drama & Theatre Studies	111	70.3	16.2	7.2	1.8	0.0	0.9	0.0	0.0	1.8	0.0	1.8
Economics	789	52.1	20.9	7.0	3.0	3.2	1.4	1.9	2.5	2.9	2.8	2.3
Electronics	10	50.0	10.0	20.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
English Language	331	71.0	20.5	4.5	0.9	0.3	0.9	0.0	0.0	0.0	0.6	1.2
English Language & Literature	144	63.2	22.9	4.9	4.9	0.7	0.0	0.7	0.7	0.0	0.0	2.1
English Literature	974	65.7	19.6	5.0	2.0	1.0	1.1	0.6	0.9	2.0	0.8	1.2
Environmental Studies	20	30.0	40.0	20.0	10.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Film Studies	111	45.9	27.9	17.1	2.7	3.6	0.0	1.8	0.0	0.0	0.0	0.9
French	70	48.6	21.4	12.9	0.0	0.0	2.9	0.0	2.9	2.9	4.3	4.3
Further Mathematics	216	48.1	15.7	5.1	3.7	2.8	5.6	3.7	3.2	4.2	3.7	4.2
Geography	623	77.2	12.8	2.2	1.4	0.8	1.1	0.5	0.5	1.3	0.8	1.3
Geology	23	47.8	34.8	8.7	4.3	0.0	0.0	0.0	0.0	4.3	0.0	0.0
German	7	71.4	14.3	14.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
History	993	71.1	15.6	4.1	1.6	0.9	0.8	0.9	0.9	1.5	1.2	1.3
History of Art	7	57.1	14.3	14.3	0.0	0.0	0.0	0.0	0.0	14.3	0.0	0.0
Italian	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Japanese	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Latin	2	0.0	0.0	50.0	0.0	0.0	0.0	50.0	0.0	0.0	0.0	0.0
Law	237	60.3	24.5	6.8	5.5	1.7	0.8	0.0	0.0	0.0	0.0	0.4
Mathematics	1678	51.7	19.2	8.9	4.5	2.6	1.5	2.6	2.2	2.6	2.0	2.1
Media Studies	336	63.1	19.3	10.4	1.8	2.4	0.9	0.0	0.3	0.0	0.3	1.5
Music	14	50.0	21.4	7.1	7.1	7.1	7.1	0.0	0.0	0.0	0.0	0.0
Music Technology	14	21.4	57.1	14.3	0.0	7.1	0.0	0.0	0.0	0.0	0.0	0.0
Panjabi	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Philosophy	41	34.1	29.3	22.0	7.3	4.9	0.0	0.0	0.0	0.0	2.4	0.0
Physical Education	146	72.6	15.1	4.8	1.4	0.0	0.7	0.7	0.0	0.7	0.0	4.1
Physics	808	55.9	20.3	5.9	2.8	1.9	1.6	1.4	1.9	3.5	2.2	2.6
Polish	0	-	-	-	-	-	-	-	-	-	-	-
Politics	380	51.1	22.4	7.9	3.9	2.9	1.3	1.3	1.3	3.4	2.9	1.6
Portuguese	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0
Psychology	1432	61.4	22.6	6.6	2.7	1.2	1.7	0.8	0.8	1.0	0.6	0.8
Religious Studies	282	59.9	27.3	5.7	1.1	2.5	0.4	0.7	1.4	0.7	0.0	0.4
Russian	6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
Sociology	877	57.8	23.7	9.7	3.5	2.2	1.1	0.3	0.2	0.3	0.3	0.7
Spanish	93	34.4	19.4	14.0	8.6	2.2	0.0	1.1	1.1	7.5	2.2	9.7
Statistics	23	56.5	26.1	8.7	4.3	0.0	4.3	0.0	0.0	0.0	0.0	0.0
Turkish	5	40.0	40.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
Urdu	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

Annex E: Predictive accuracy data tables

Table E.1. Data table for Figure 7.3

Approach	A Level Biology	A Level French	A Level Drama	A Level RS
Approach 1	0.657	0.500	0.529	0.567
Approach 2	0.663	0.498	0.527	0.564
Approach 3	0.664	0.515	0.538	0.584
Approach 4	0.568	0.428	0.452	0.484
Approach 5	0.554	0.422	0.439	0.471
Approach 6	0.504	0.414	0.455	0.464
Approach 7	0.656	0.513	0.540	0.584
Approach 8	0.380	0.360	0.360	0.330
Approach 9	0.653	0.512	0.539	0.582
Approach 10	0.657	0.522	0.537	0.580
Approach 11	0.621	0.503	0.507	0.580

Table E.2. Data table for Figure 7.4

Approach	GCSE English Language	GCSE Mathematics	GCSE History	GCSE Music
Approach 1	0.747	0.740	0.563	0.394
Approach 2	0.743	0.735	0.576	0.409
Approach 3	0.748	0.742	0.567	0.407
Approach 4	0.558	0.616	0.441	0.341
Approach 5	0.538	0.598	0.411	0.320
Approach 6	0.472	0.489	0.384	0.307
Approach 7	0.703	0.677	0.568	0.398
Approach 8	0.353	0.380	0.215	0.228
Approach 9	0.727	0.714	0.560	0.401
Approach 11	0.651	0.627	0.509	0.368

Table E.3. Data table for Figure 7.5.

	Number of centres with the following numbers of candidates						
	>=500	250-499	100-249	50-99	30-49	10-29	1-9
GCSE English language	82	441	2,636	561	252	682	697

Table E.4. Data table for Figure 7.6.

	Number of centres with the following numbers of candidates					
	>=100	**75-99**	**50-74**	25-49	**10-24**	**1-9**
A level biology	52	49	150	608	982	736

Table E.5. Data table for Figures 7.7 and 7.8

		Number of centres with 2018 outcomes in the following intervals				
GCSE English language	Grade 7	**>=75%**	**[50%,75%)**	**[20%,50%)**	**[5%,20%)**	**<5%**
		35	158	973	2,205	1731
	Grade 4	**>=90%**	**[75%,90%)**	**[50%,75%)**	**[20%,50%)**	**<20%**
		593	972	2,027	932	578

Table E.6. Data table for Figures 7.9 and 7.10.

		Number of centres with 2018 outcomes in the following intervals				
A level biology	Grade A	**>=75%**	**[50%,75%)**	**[25%,50%)**	**[10%,25%)**	**<10%**
		33	178	664	788	808
	Grade C	**>=75%**	**[50%,75%)**	**[25%,50%)**	**[10%,25%)**	**<10%**
		922	920	424	99	106

Table E.7. Data table for Figures 7.11 and 7.12.

	Number of centres with the following prior attainment match rate					
	>=90%	**[75%,90%)**	**[50%,75%)**	**[25%,50%)**	**[10%,25%)**	**<10%**
GCSE English language	2,049	1,568	613	299	86	137
A level biology	1,857	231	131	74	37	167

Table E.8. Data tables for Figures 7.19 and 7.22

Subject	**Number of candidates**	**Predictive Accuracy**	**Accuracy within a grade**
History	47044	0.679	0.986
English Language	13560	0.659	0.987
Psychology	61882	0.652	0.985
Biology	63580	0.650	0.981
Ancient History	681	0.643	0.979
Economics	29740	0.642	0.977
English Language & Literature	7425	0.641	0.980
English Literature	36913	0.640	0.979
Sociology	35610	0.637	0.979
Chemistry	54868	0.632	0.970
Business Studies	30236	0.625	0.979
Media Studies	14004	0.625	0.981
Russian	695	0.620	0.944
Geography	31279	0.619	0.980
Chinese	2086	0.607	0.963
Mathematics	83898	0.606	0.961
Art & Design	38597	0.596	0.960
Film Studies	5765	0.592	0.973
Environmental Studies	877	0.592	0.954
Classical Greek	210	0.590	0.910
Physics	36025	0.585	0.955
Religious Studies	16220	0.565	0.961
Politics	18091	0.562	0.960
Law	11092	0.555	0.956
Accounting	2215	0.552	0.928
Classical Civilisation	2893	0.551	0.960
Philosophy	2768	0.546	0.945
Geology	1157	0.541	0.956
Statistics	669	0.534	0.942
Latin	1103	0.526	0.915
Drama & Theatre Studies	8970	0.522	0.948
Electronics	588	0.519	0.944

Panjabi	182	0.500	0.901
Further Mathematics	13473	0.499	0.886
Computing	10435	0.498	0.927
French	7589	0.497	0.917
Sport & PE	9832	0.493	0.934
Dutch	33	0.485	0.788
Biblical Hebrew	66	0.485	0.939
Spanish	7786	0.479	0.909
Polish	1004	0.472	0.842
Music	5056	0.470	0.919
D&T: Product Design	9214	0.466	0.923
Modern Hebrew	52	0.462	0.769
Dance	1016	0.435	0.923
German	2857	0.421	0.870
Turkish	581	0.411	0.826
Japanese	195	0.410	0.887
Urdu	401	0.374	0.863
Greek (Modern)	186	0.360	0.796
Portuguese	494	0.346	0.800
Gujarati	25	0.320	0.640
Bengali	28	0.286	0.786
Arabic	765	0.281	0.722
Italian	799	0.267	0.750

Table E.9. Data tables for Figures 7.21 and 7.24

Subject	**Number of candidates**	**Predictive Accuracy**	**Accuracy within a grade**
English Language	706620	0.745	0.993
Mathematics	719874	0.735	0.992
English Literature	546638	0.700	0.990
Geography	250986	0.596	0.975
Classical Greek	1143	0.590	0.885
Biology	165072	0.578	0.947
Physics	157787	0.574	0.950
Latin	9355	0.562	0.886
Chemistry	159038	0.558	0.941
Religious Studies	249325	0.558	0.951
History	260468	0.552	0.961
French	121756	0.483	0.898
German	40655	0.479	0.895
Citizenship Studies	19644	0.478	0.909
Art & Design	182216	0.472	0.914
Physical Education	79942	0.463	0.905
Food Preparation and Nutrition	44921	0.456	0.900

Spanish	95860	0.445	0.858
Computing	77419	0.432	0.883
Drama	57560	0.428	0.887
Music	34049	0.393	0.825
Dance	9274	0.337	0.775

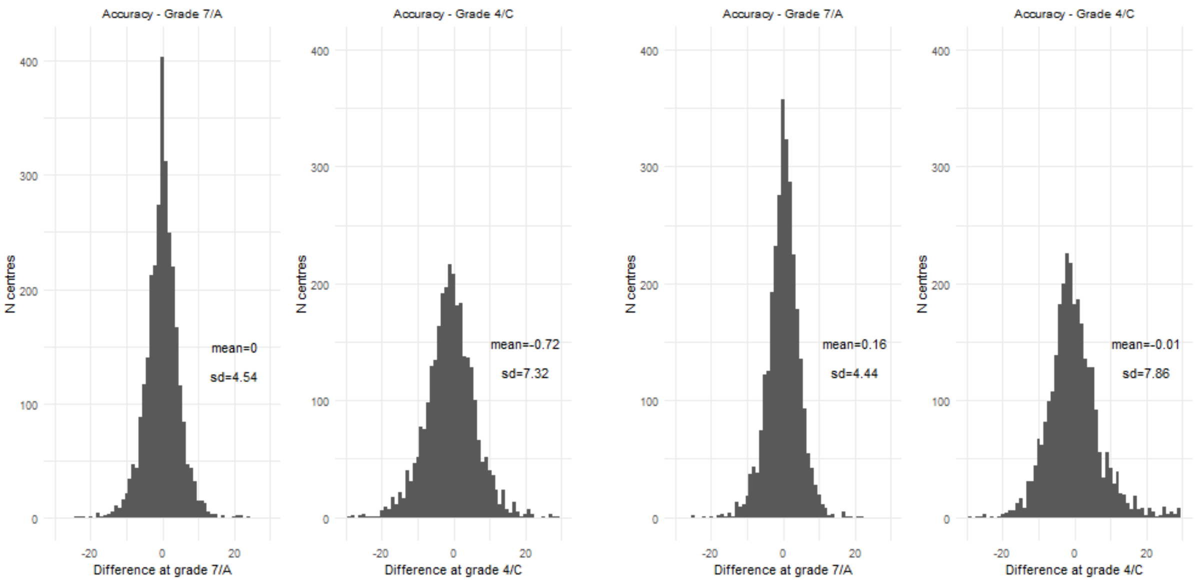
Annex F: Accuracy of predictions

GCSE English language

Figure F.1. Univariate distribution of key variables – English language

a) Distribution of accuracy at centre level

DCP *DCP Modified*



b) Distribution of demographic and socio-economic centre composition

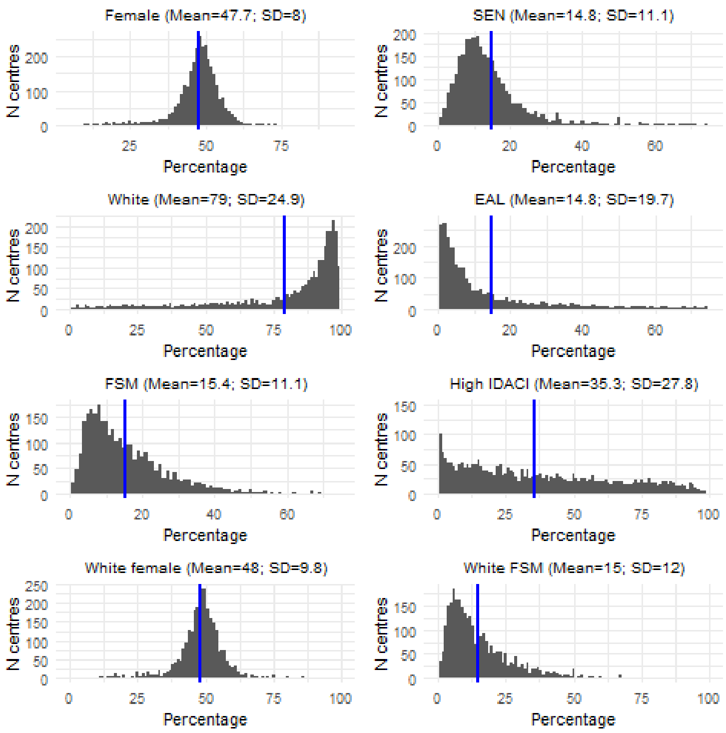


Figure F.2. Correlation between demographic/socio-economic factors and accuracy of predictions – English language

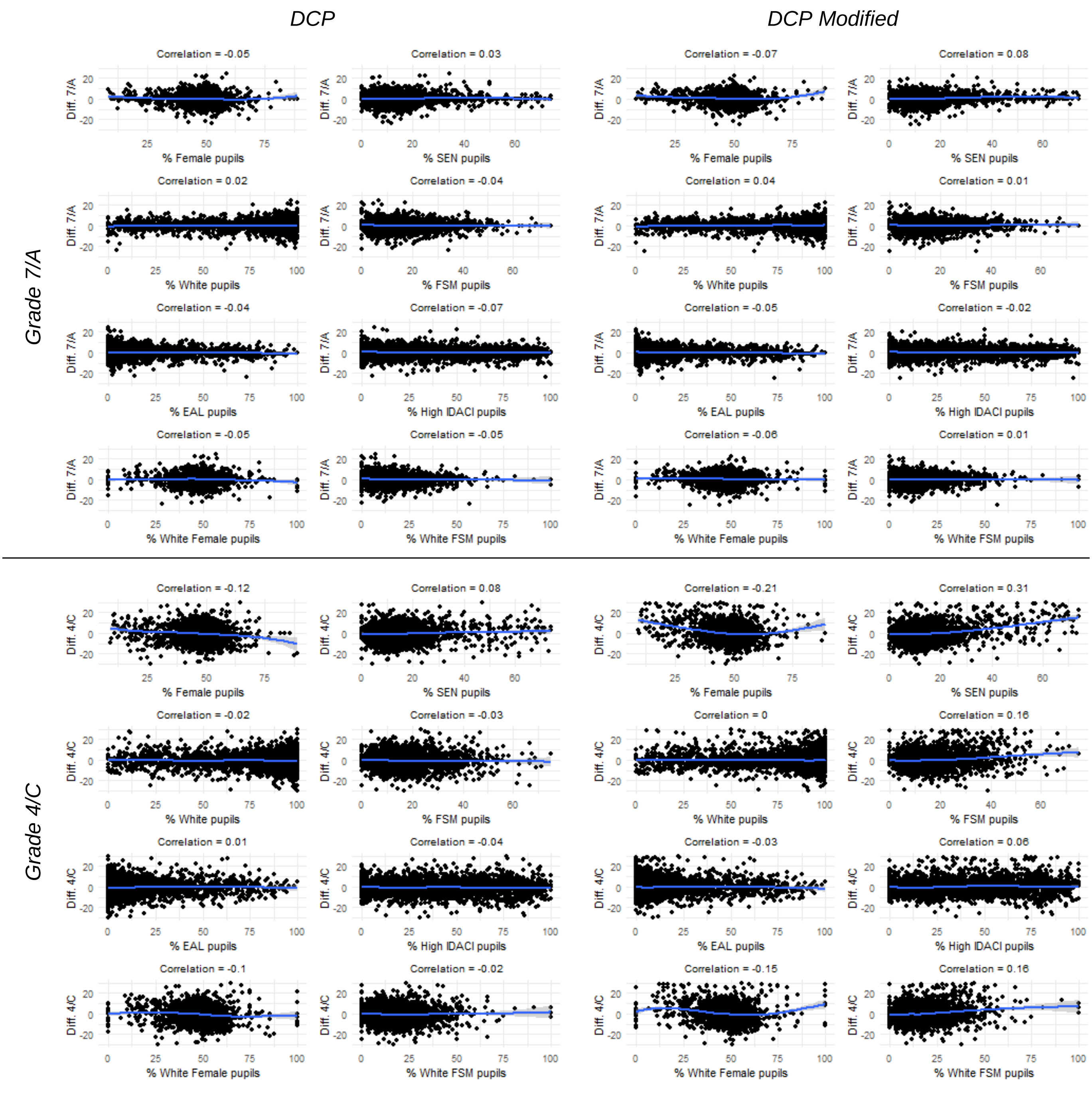


Table F.1. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – English language

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	0.07	-0.77 *	0.49 **	1.66 ***
	(0.19)	(0.30)	(0.18)	(0.31)
Perc.Female	-0.02	-0.08 **	-0.03	-0.14 ***
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.SEN	0.02	0.05 ***	0.02 **	0.15 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.WHITE	-0.01	-0.02	-0.01	-0.04 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.FSM	0.01	-0.06	0.01	-0.01
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.EAL	-0.01	-0.00	-0.02 *	-0.05 ***
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.high	-0.01 *	-0.02 *	-0.00	-0.02 **
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.med	0.00	-0.00	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
WHITE_Female	-0.00	-0.01	-0.01	0.01
	(0.01)	(0.02)	(0.01)	(0.02)
WHITE_FSM	-0.02	0.02	-0.01	0.05
	(0.02)	(0.03)	(0.02)	(0.03)
mean.ks2	-0.24	-1.87 *	-0.56	-3.49 ***
	(0.55)	(0.88)	(0.53)	(0.88)
Ncands_2019	-0.00	0.00	-0.00 *	-0.01 ***
	(0.00)	(0.00)	(0.00)	(0.00)
N	2981	2981	2981	2981
R2	0.01	0.03	0.02	0.14

***** p < 0.001; ** p < 0.01; * p < 0.05.**

Table F.2. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – English language

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	0.06	-0.81 **	0.48 **	1.65 ***
	(0.19)	(0.30)	(0.18)	(0.31)
Perc.Female	-0.02	-0.08 **	-0.03	-0.14 ***
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.SEN	0.02	0.05 ***	0.02 **	0.15 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.AOEG	0.02	-0.04	0.02	-0.04
	(0.03)	(0.05)	(0.03)	(0.05)
Perc.ASIA	0.01	-0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.BLAC	0.02	0.04 *	0.01	0.04 *
	(0.01)	(0.02)	(0.01)	(0.02)
Perc.CHIN	0.15	0.12	0.11	0.13
	(0.13)	(0.21)	(0.13)	(0.22)
Perc.MIXD	0.00	0.04	0.00	0.12 ***
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.FSM	0.01	-0.06	0.00	-0.01
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.EAL	-0.01	0.01	-0.02 *	-0.02
	(0.01)	(0.02)	(0.01)	(0.02)
IDACI.high	-0.01 *	-0.02 *	-0.00	-0.03 **
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.med	0.00	-0.00	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
WHITE_Female	-0.00	-0.01	-0.00	0.01
	(0.01)	(0.02)	(0.01)	(0.02)
WHITE_FSM	-0.02	0.03	-0.01	0.05
	(0.02)	(0.03)	(0.02)	(0.03)
mean.ks2	-0.38	-1.83 *	-0.67	-3.38 ***
	(0.56)	(0.90)	(0.55)	(0.91)

Ncands_2019	-0.00	0.00	-0.00	-0.01 ***
	(0.00)	(0.00)	(0.00)	(0.00)
N	2981	2981	2981	2981
R2	0.01	0.03	0.02	0.14

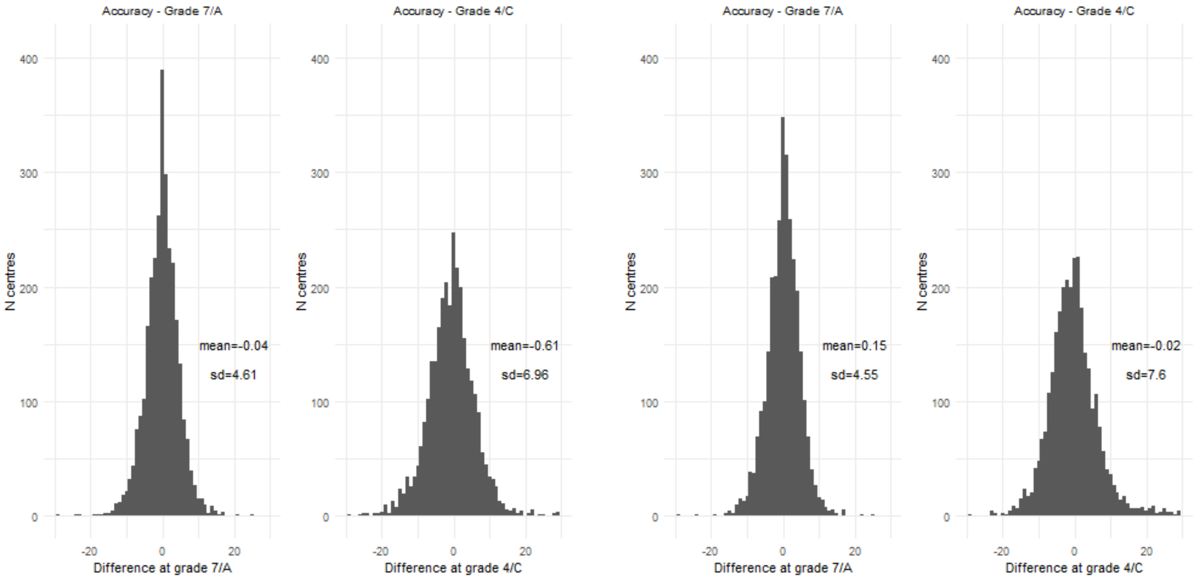
***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE Mathematics

Figure F.3. Univariate distribution of key variables – Maths

a. Distribution of accuracy at centre level

DCP | *DCP Modified*



b. Distribution of demographic and socio-economic centre composition

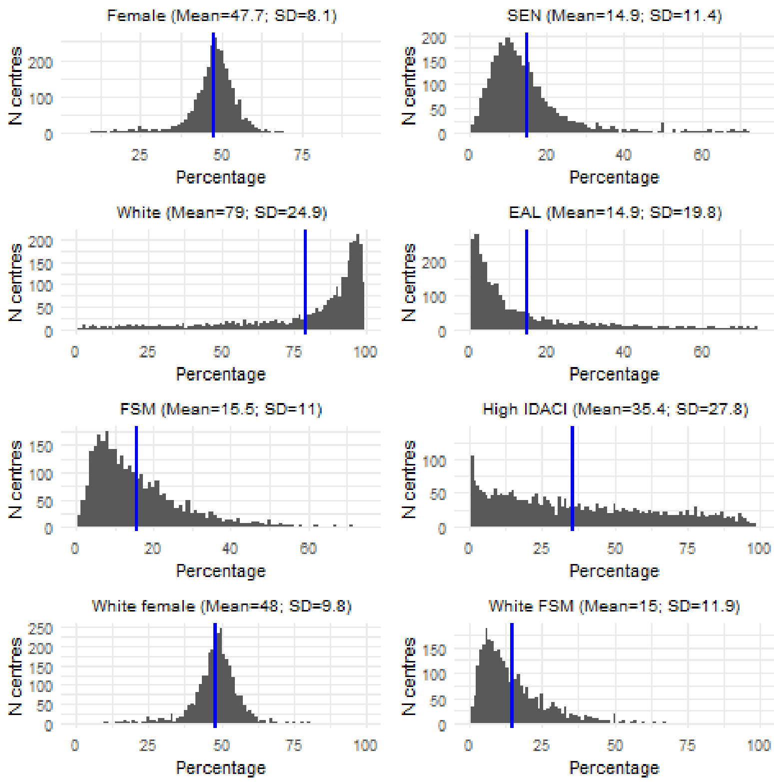


Figure F.4. Correlation between demographic/socio-economic factors and accuracy of predictions – Maths

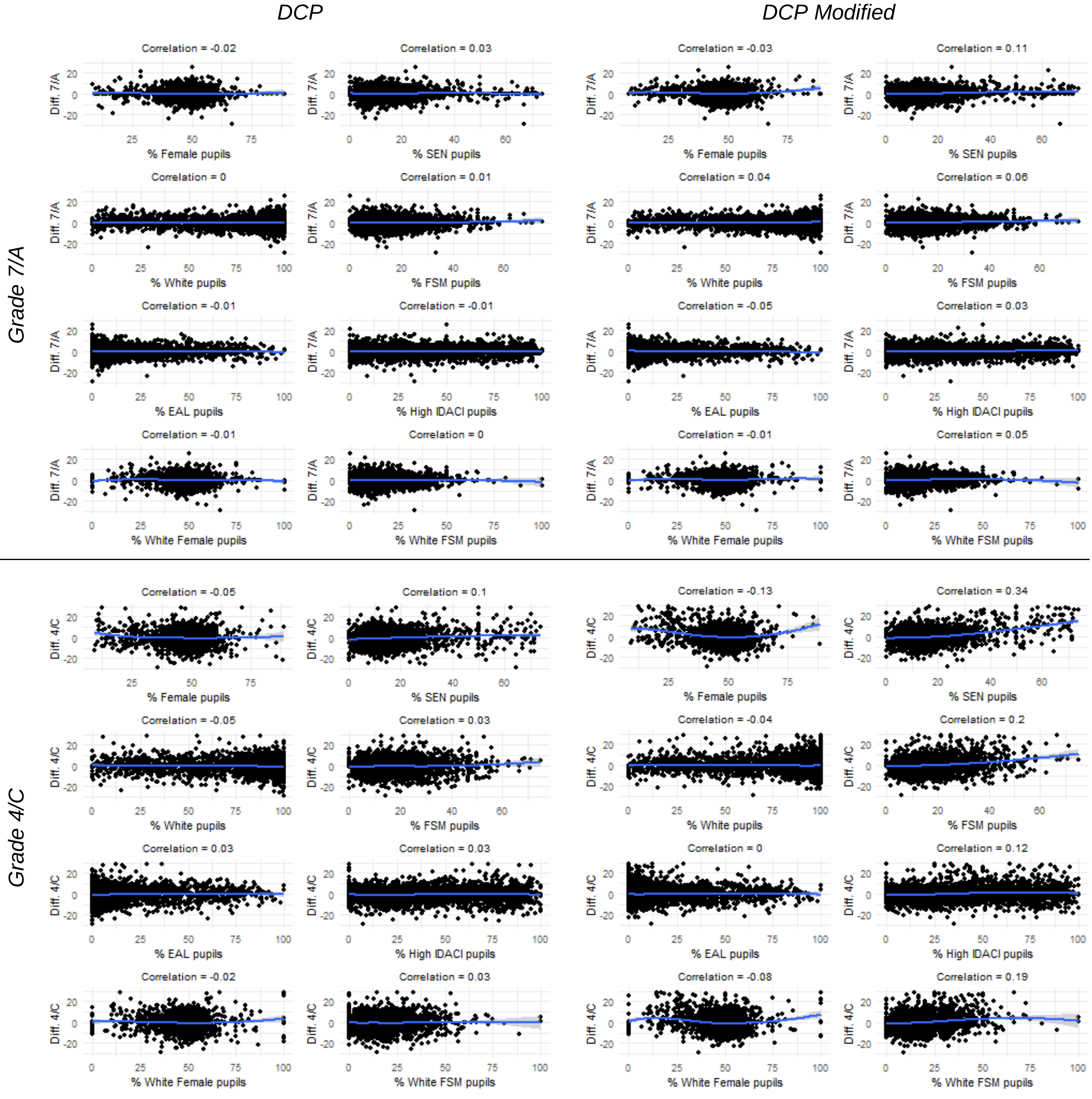


Table F.3. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Maths

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	-0.04	-1.16 ***	0.41 *	0.87 **
	(0.19)	(0.28)	(0.18)	(0.29)
Perc.Female	-0.01	-0.07 **	-0.02	-0.09 ***
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.SEN	0.01	0.07 ***	0.03 ***	0.19 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.WHITE	-0.01	-0.02	0.00	-0.04 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.FSM	0.02	-0.02	0.02	0.03
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.EAL	-0.01	-0.00	-0.01	-0.05 ***
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.high	-0.00	0.00	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.med	0.00	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
WHITE_Female	0.00	0.05 *	0.01	0.05 *
	(0.01)	(0.02)	(0.01)	(0.02)
WHITE_FSM	-0.02	-0.00	-0.02	0.01
	(0.02)	(0.03)	(0.02)	(0.03)
mean.ks2	-0.46	-0.59	-0.68	-1.34
	(0.50)	(0.74)	(0.48)	(0.76)
Ncands_2019	0.00	0.00 *	-0.00	-0.01 ***
	(0.00)	(0.00)	(0.00)	(0.00)
N	2992	2992	2992	2992
R2	0.00	0.02	0.02	0.14

***** p < 0.001; ** p < 0.01; * p < 0.05.**

Table F.4. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Maths

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	-0.07	-1.20 ***	0.38 *	0.85 **
	(0.19)	(0.28)	(0.18)	(0.29)
Perc.Female	-0.01	-0.07 **	-0.01	-0.09 **
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.SEN	0.01	0.07 ***	0.03 ***	0.19 ***
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.AOEG	-0.01	-0.01	-0.02	0.03
	(0.03)	(0.04)	(0.03)	(0.04)
Perc.ASIA	-0.01	0.00	-0.01	0.03
	(0.01)	(0.01)	(0.01)	(0.01)
Perc.BLAC	0.02	0.05 **	0.01	0.05 **
	(0.01)	(0.02)	(0.01)	(0.02)
Perc.CHIN	0.04	0.07	-0.00	0.05
	(0.14)	(0.21)	(0.13)	(0.21)
Perc.MIXD	0.01	-0.01	0.00	0.07 *
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.FSM	0.02	-0.02	0.02	0.03
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.EAL	-0.00	0.00	-0.01	-0.04 **
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.high	-0.01	-0.00	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.med	0.00	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
WHITE_Female	0.00	0.05 *	0.01	0.05 *
	(0.01)	(0.02)	(0.01)	(0.02)
WHITE_FSM	-0.01	0.00	-0.02	0.02
	(0.02)	(0.03)	(0.02)	(0.03)
mean.ks2	-0.48	-0.64	-0.67	-1.31
	(0.50)	(0.75)	(0.49)	(0.77)
Ncands_2019	0.00	0.00 *	-0.00	-0.01 ***
	(0.00)	(0.00)	(0.00)	(0.00)

N	2992	2992	2992	2992
R2	0.00	0.02	0.02	0.14

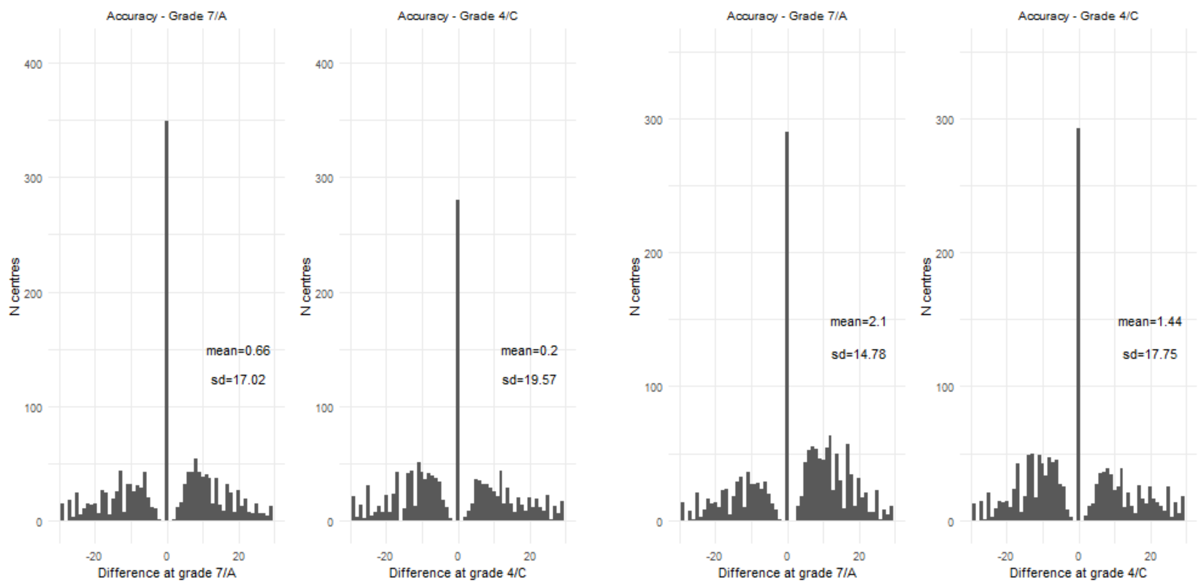
***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE Music

Figure F.5. Univariate distribution of key variables – Music

a. Distribution of accuracy at centre level

DCP *DCP Modified*



b. Distribution of demographic and socio-economic centre composition

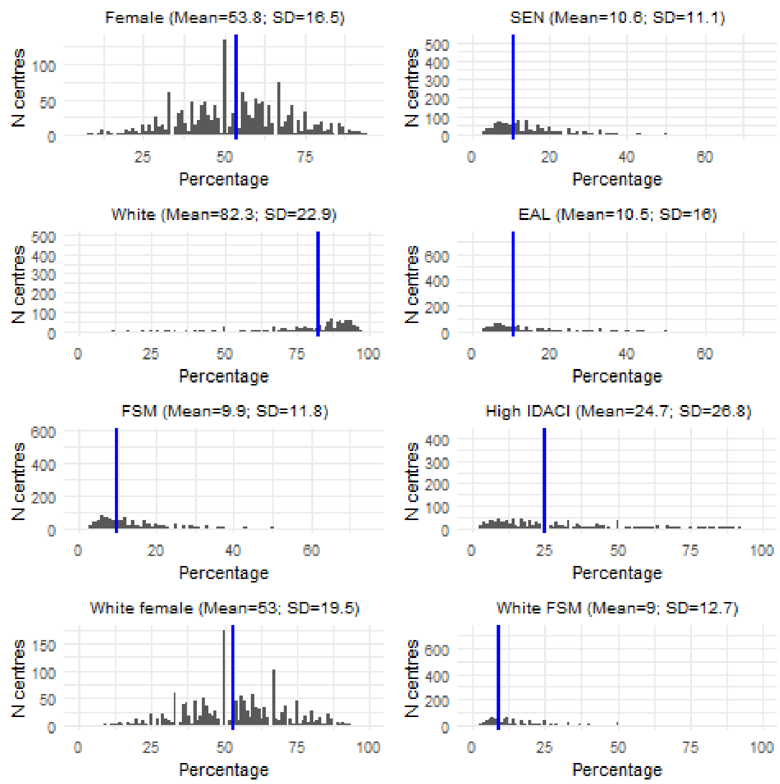


Figure F.6. Correlation between demographic/socio-economic factors and accuracy of predictions – Music

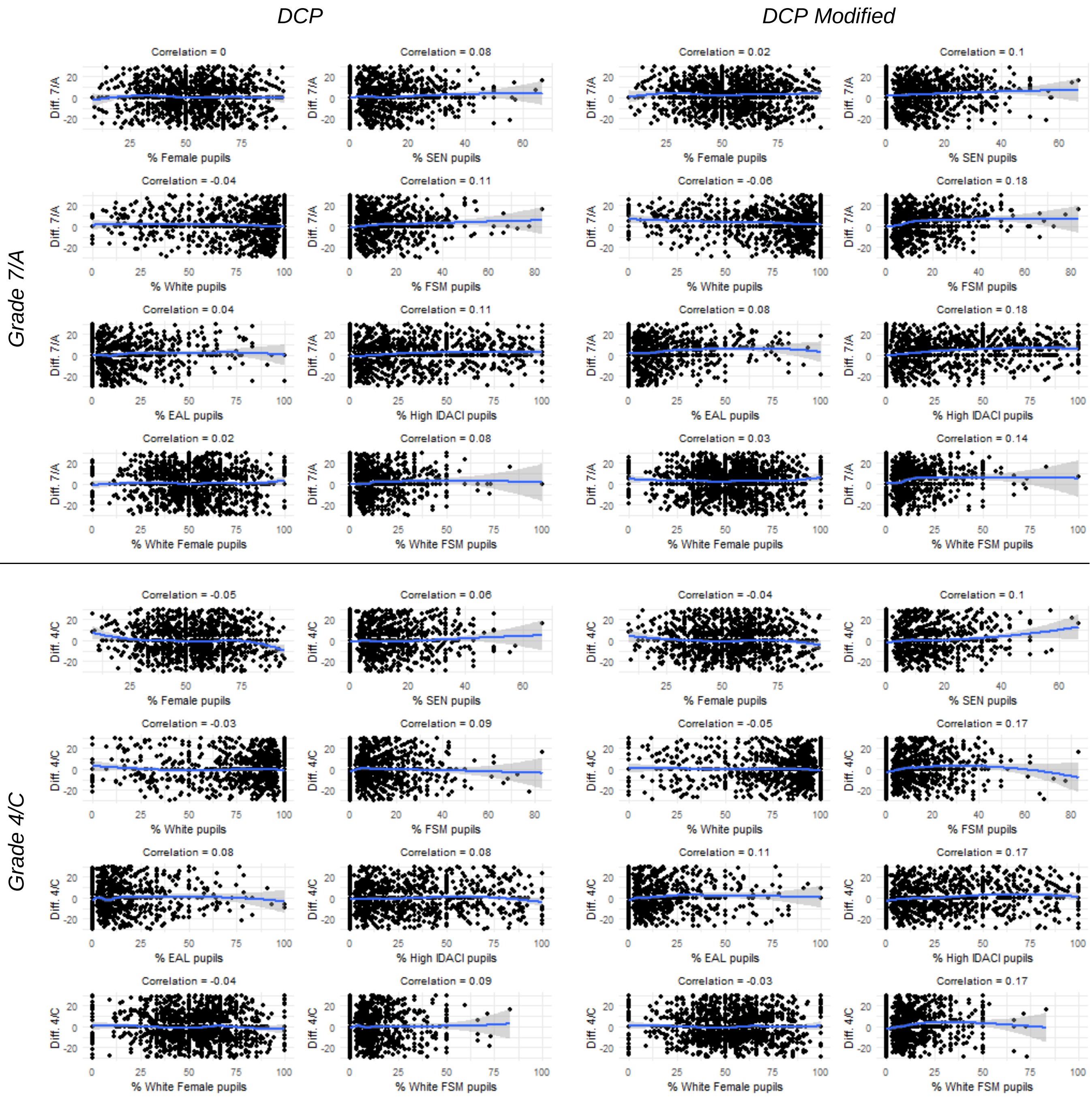


Table F.5. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Music

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	-0.38	-1.43	2.27 **	1.30
	(0.99)	(1.13)	(0.84)	(1.01)
Perc.Female	-0.05	-0.06	-0.03	-0.03
	(0.05)	(0.06)	(0.04)	(0.05)
Perc.SEN	0.05	0.00	0.05	0.02
	(0.04)	(0.05)	(0.04)	(0.04)
Perc.WHITE	0.01	0.04	0.03	0.08 **
	(0.03)	(0.03)	(0.02)	(0.03)
Perc.FSM	0.14	0.01	0.17 **	0.02
	(0.07)	(0.08)	(0.06)	(0.07)
Perc.EAL	-0.00	0.10 *	0.01	0.12 **
	(0.04)	(0.04)	(0.03)	(0.04)
IDACI.high	0.05 *	0.01	0.08 ***	0.07 **
	(0.02)	(0.03)	(0.02)	(0.02)
IDACI.med	0.04	-0.01	0.08 ***	0.05 *
	(0.02)	(0.03)	(0.02)	(0.02)
WHITE_Female	0.06	0.01	0.04	-0.01
	(0.04)	(0.05)	(0.04)	(0.04)
WHITE_FSM	-0.10	0.06	-0.09	0.09
	(0.06)	(0.07)	(0.05)	(0.06)
mean.ks2	-4.23 *	-7.12 ***	-4.09 **	-6.63 ***
	(1.84)	(2.12)	(1.57)	(1.89)
Ncands_2019	0.07	0.10	-0.01	0.01
	(0.06)	(0.07)	(0.05)	(0.06)
N	1595	1595	1595	1595
R2	0.03	0.03	0.07	0.06

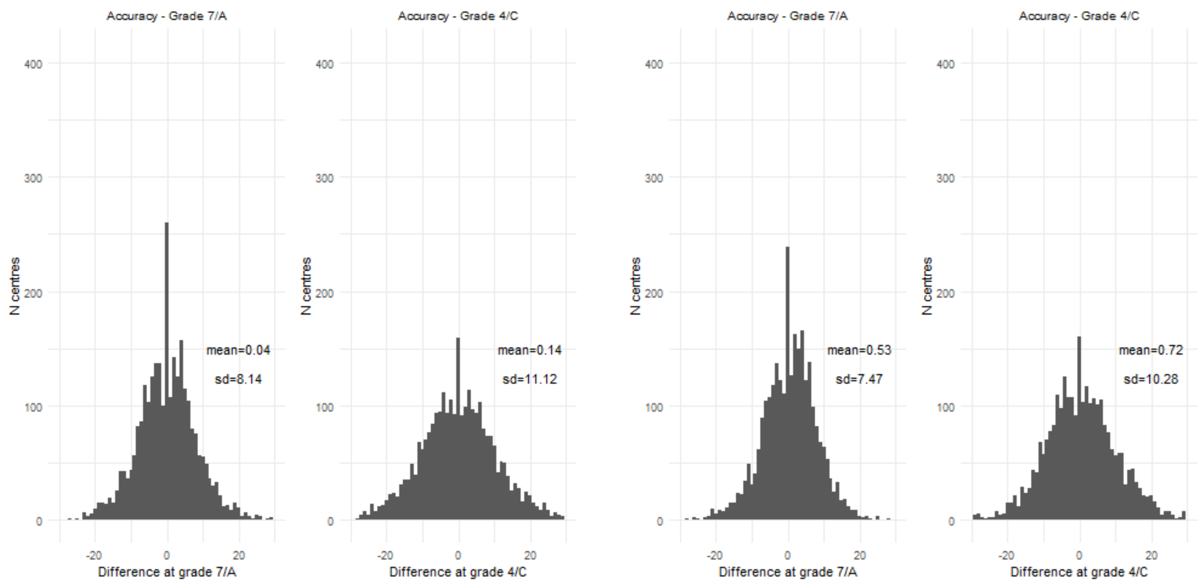
***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE History

Figure F.7. Univariate distribution of key variables – History

a. Distribution of accuracy at centre level

DCP *DCP Modified*



b. Distribution of demographic and socio-economic centre composition

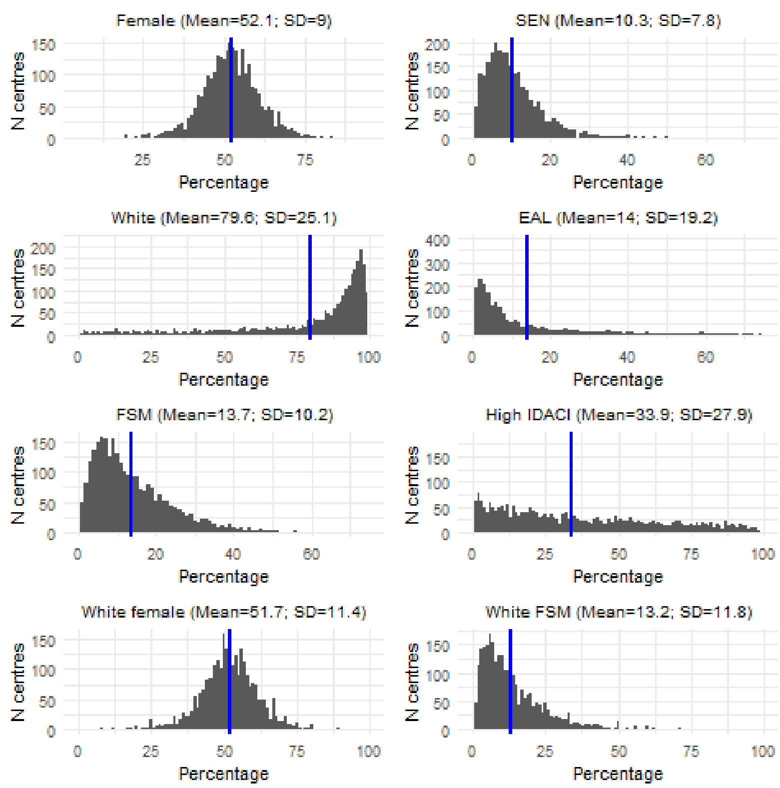


Figure F.8. Correlation between demographic/socio-economic factors and accuracy of predictions – History

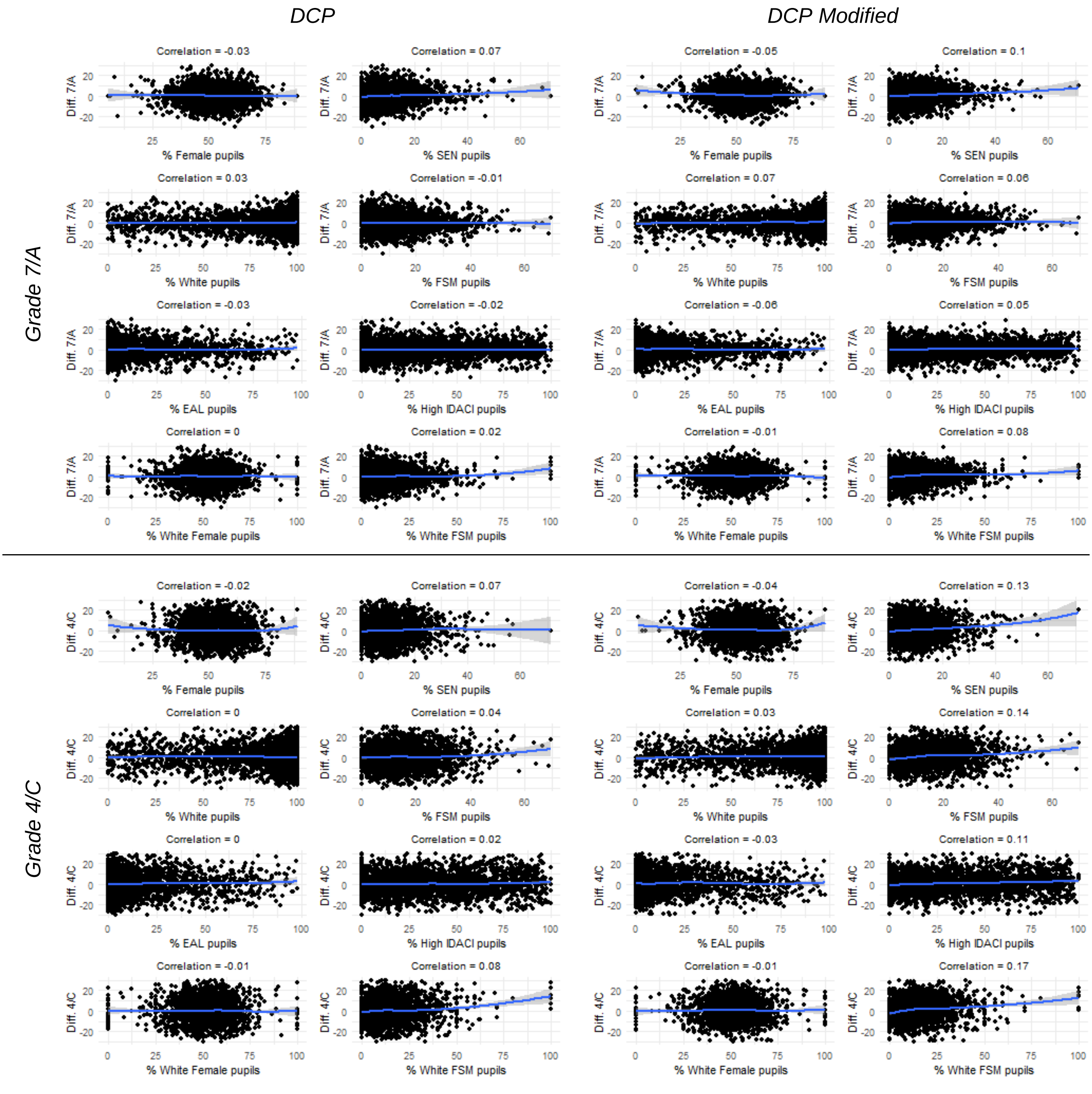


Table F.6. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – History

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	0.29	0.42	1.04 **	1.32 **
	(0.38)	(0.52)	(0.34)	(0.47)
Perc.Female	-0.06 *	-0.05	-0.07 **	-0.09 **
	(0.03)	(0.04)	(0.02)	(0.03)
Perc.SEN	0.07 **	0.07 *	0.06 **	0.10 ***
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.WHITE	0.00	-0.02	0.02 *	0.02
	(0.01)	(0.02)	(0.01)	(0.02)
Perc.FSM	-0.07 *	-0.14 **	-0.04	-0.07
	(0.04)	(0.05)	(0.03)	(0.04)
Perc.EAL	-0.00	-0.02	-0.01	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)
IDACI.high	-0.01	-0.01	0.02 *	0.04 **
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.med	-0.01	-0.01	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)
WHITE_Female	0.04	0.03	0.03	0.03
	(0.02)	(0.03)	(0.02)	(0.03)
WHITE_FSM	0.07 **	0.16 ***	0.05 *	0.14 ***
	(0.02)	(0.03)	(0.02)	(0.03)
mean.ks2	-0.98	-1.96	-1.27	-2.42 *
	(0.94)	(1.28)	(0.85)	(1.16)
Ncands_2019	-0.00	-0.00	-0.01	-0.01
	(0.00)	(0.01)	(0.00)	(0.01)
N	2760	2760	2760	2760
R2	0.01	0.02	0.03	0.05

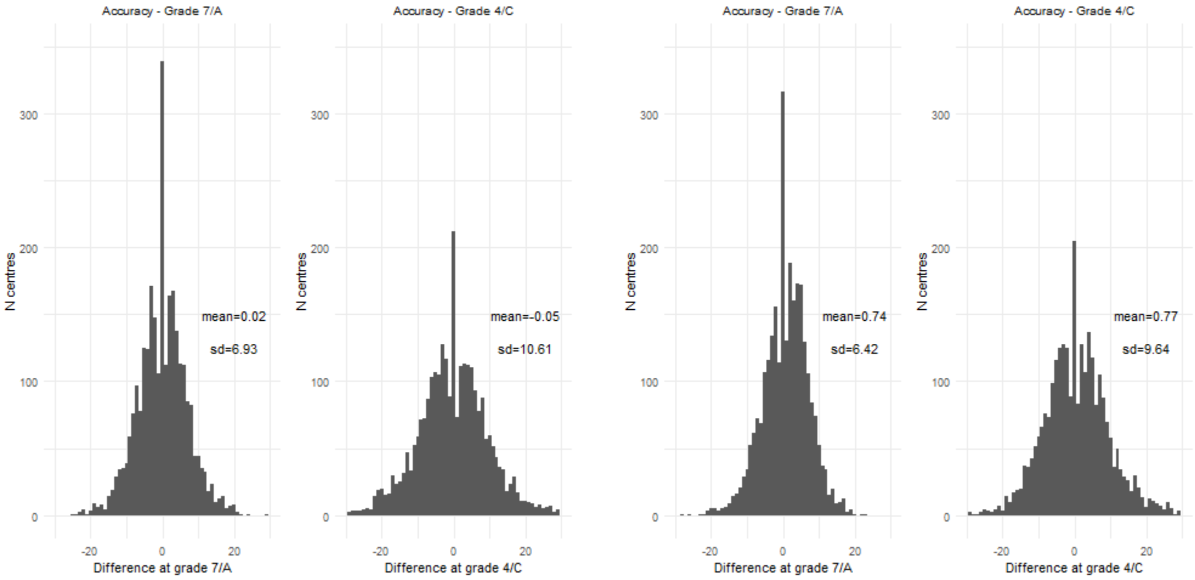
***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE Geography

Figure F.9. Univariate distribution of key variables – Geography

a. Distribution of accuracy at centre level

DCP | *DCP Modified*



b. Distribution of demographic and socio-economic centre composition

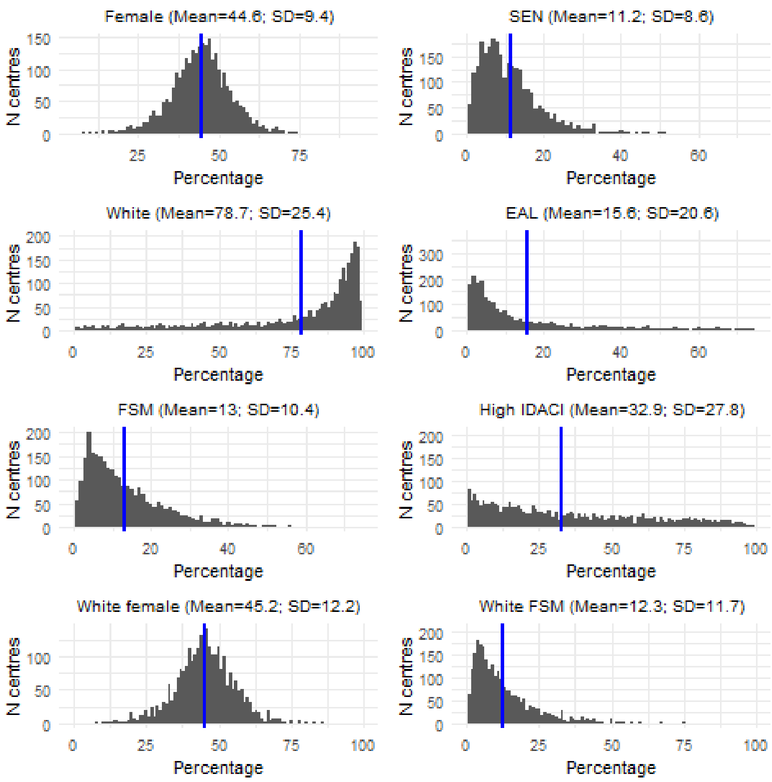


Figure F.10. Correlation between demographic/socio-economic factors and accuracy of predictions – Geography

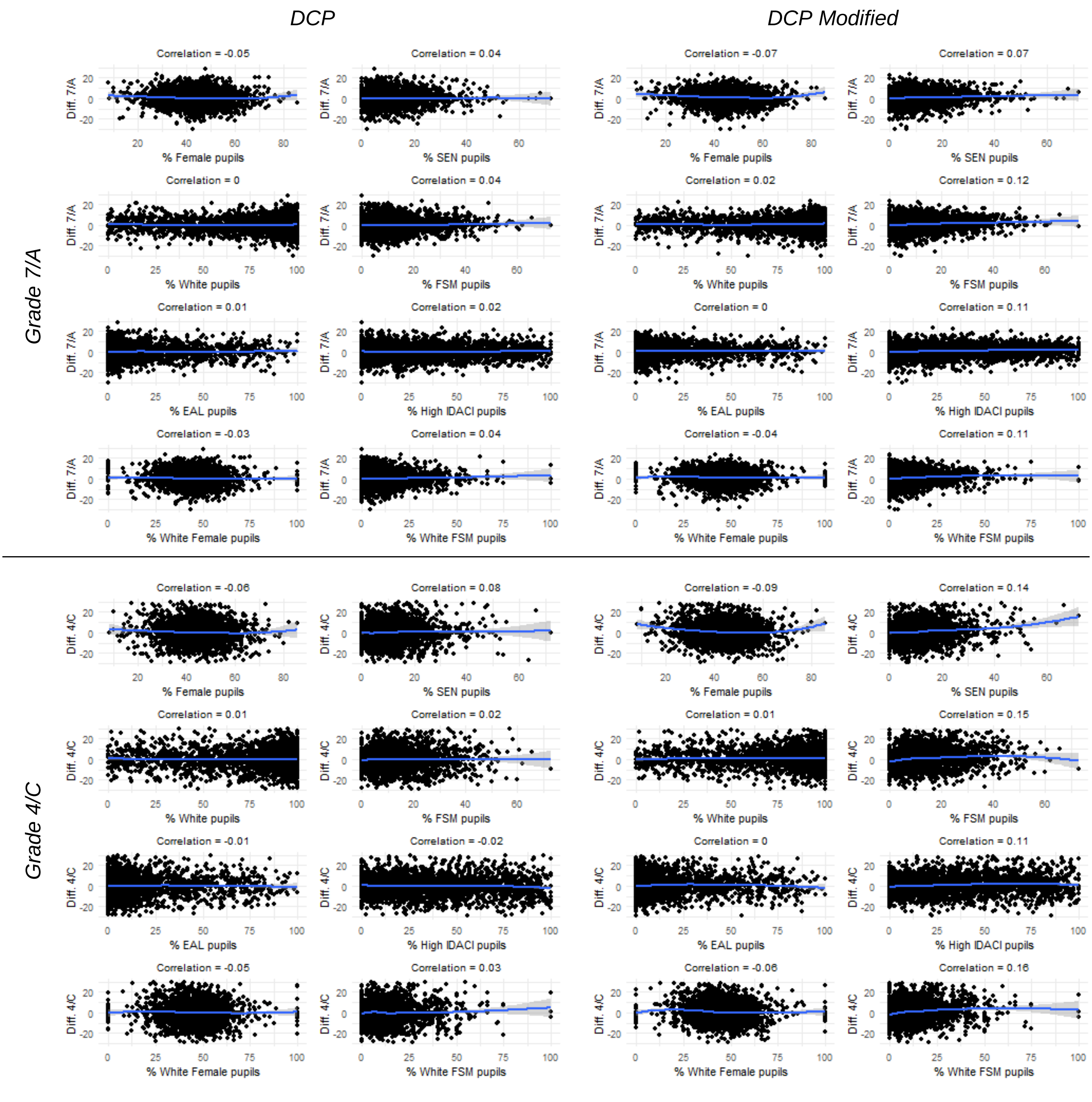


Table F.7. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Geography

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	-0.16	-0.78	1.09 ***	0.88 *
	(0.32)	(0.49)	(0.29)	(0.44)
Perc.Female	-0.05 *	-0.06	-0.06 **	-0.08 **
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.SEN	0.02	0.08 **	0.02	0.09 ***
	(0.02)	(0.03)	(0.02)	(0.02)
Perc.WHITE	0.01	-0.01	0.03 **	0.02
	(0.01)	(0.02)	(0.01)	(0.02)
Perc.FSM	0.01	-0.00	0.02	0.04
	(0.03)	(0.04)	(0.03)	(0.04)
Perc.EAL	0.01	-0.00	0.00	-0.02
	(0.01)	(0.02)	(0.01)	(0.02)
IDACI.high	0.00	-0.04 **	0.03 ***	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
IDACI.med	0.01	-0.01	0.03 ***	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
WHITE_Female	0.01	-0.00	0.01	-0.00
	(0.02)	(0.02)	(0.01)	(0.02)
WHITE_FSM	0.01	0.05	0.01	0.05
	(0.02)	(0.03)	(0.02)	(0.03)
mean.ks2	-0.83	-3.82 **	-0.72	-3.44 **
	(0.78)	(1.18)	(0.71)	(1.06)
Ncands_2019	0.00	0.01	-0.00	-0.00
	(0.00)	(0.01)	(0.00)	(0.01)
N	2768	2768	2768	2768
R2	0.01	0.02	0.04	0.05

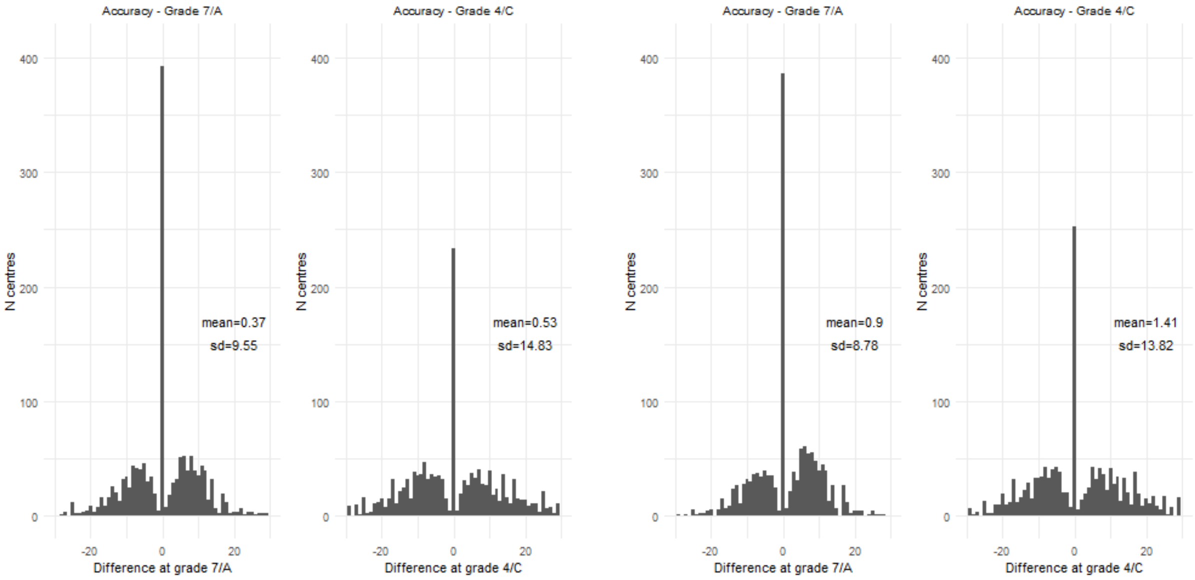
***** p < 0.001; ** p < 0.01; * p < 0.05.**

A level Biology

Figure F.11. Univariate distribution of key variables – Biology

a. Distribution of accuracy at centre level

DCP *DCP Modified*



b. Distribution of demographic and socio-economic centre composition

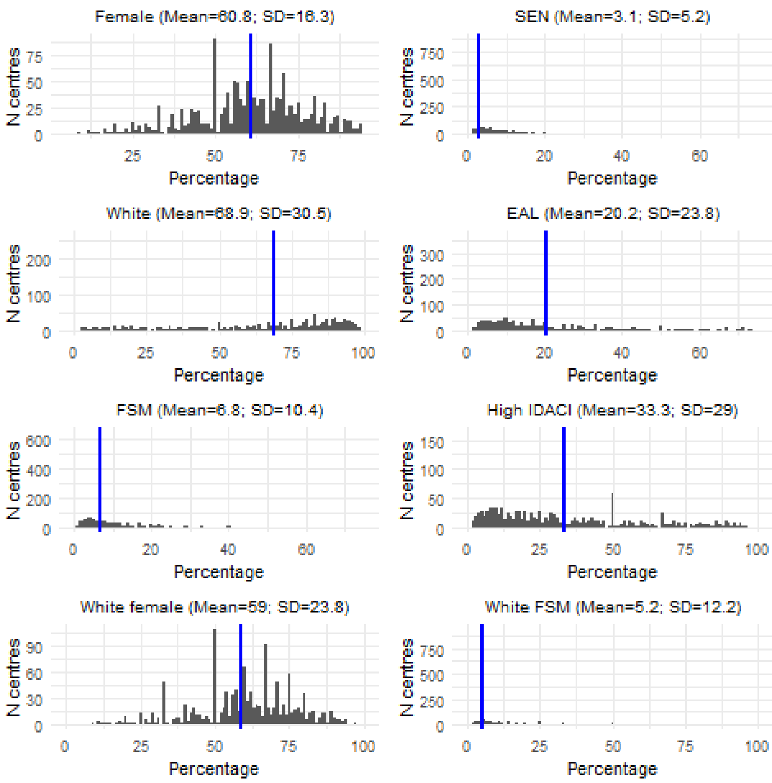


Figure F.12. Correlation between demographic/socio-economic factors and accuracy of predictions – Biology

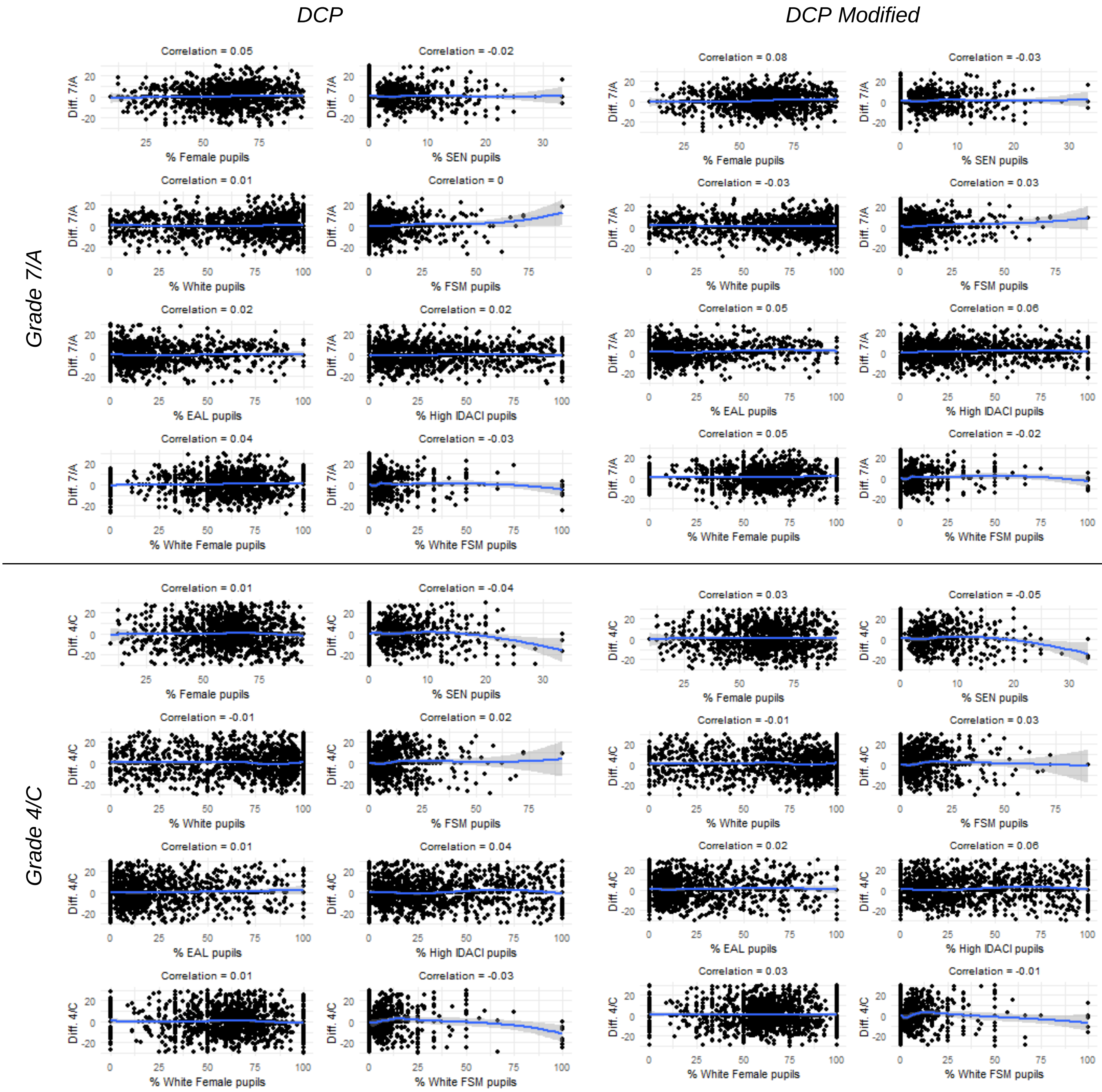


Table F.8. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Biology

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	8.25 *	51.87 ***	4.81	54.04 ***
	(3.49)	(5.26)	(3.20)	(4.85)
Perc.Female	0.03	0.01	0.04	0.03
	(0.02)	(0.03)	(0.02)	(0.03)
Perc.SEN	-0.02	-0.12	-0.03	-0.14 *
	(0.05)	(0.07)	(0.05)	(0.07)
Perc.WHITE	0.02	0.00	0.01	-0.00
	(0.02)	(0.02)	(0.01)	(0.02)
Perc.FSM	0.01	0.06	0.02	0.03
	(0.03)	(0.05)	(0.03)	(0.05)
Perc.EAL	0.02	-0.02	0.02	-0.02
	(0.02)	(0.03)	(0.02)	(0.03)
IDACI.high	0.01	0.01	0.02	0.03
	(0.02)	(0.02)	(0.01)	(0.02)
IDACI.med	0.01	-0.01	0.01	0.01
	(0.02)	(0.02)	(0.01)	(0.02)
WHITE_Female	0.01	0.03	0.01	0.03
	(0.01)	(0.02)	(0.01)	(0.02)
WHITE_FSM	-0.04	-0.10 **	-0.04	-0.07 *
	(0.02)	(0.04)	(0.02)	(0.03)
mean.GCSE	-0.12 *	-0.80 ***	-0.05	-0.80 ***
	(0.05)	(0.08)	(0.05)	(0.07)
Ncands_2019	0.01	0.07 **	-0.01	0.03
	(0.02)	(0.02)	(0.01)	(0.02)
N	1430	1430	1430	1430
R2	0.01	0.07	0.02	0.09

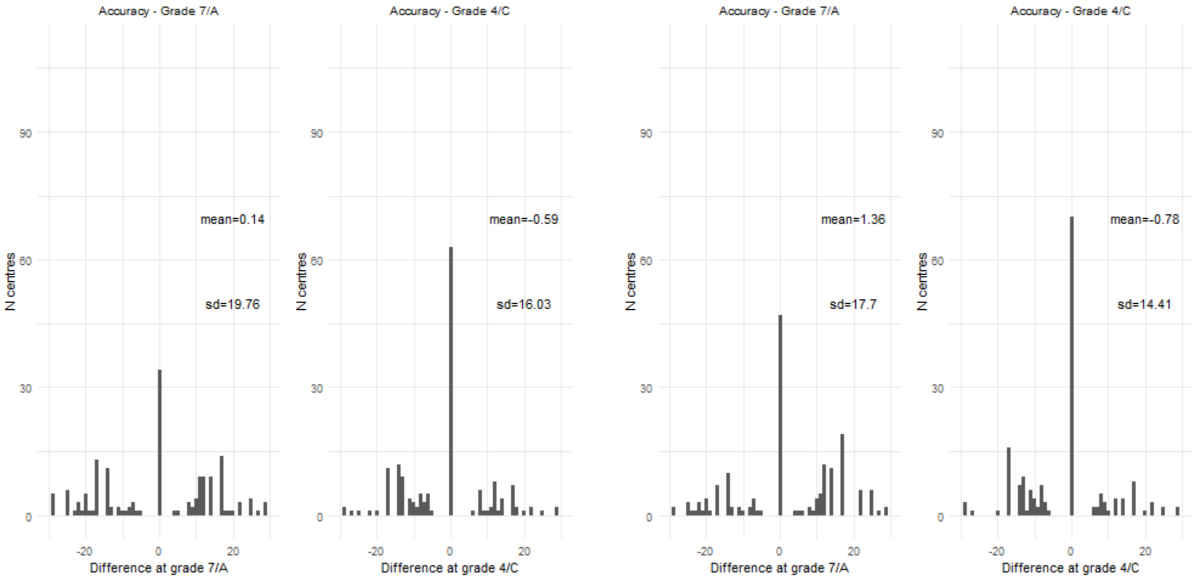
***** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.**

A level French

Figure F.13. Univariate distribution of key variables – French

a. Distribution of accuracy at centre level

DCP DCP Modified



b. Distribution of demographic and socio-economic centre composition

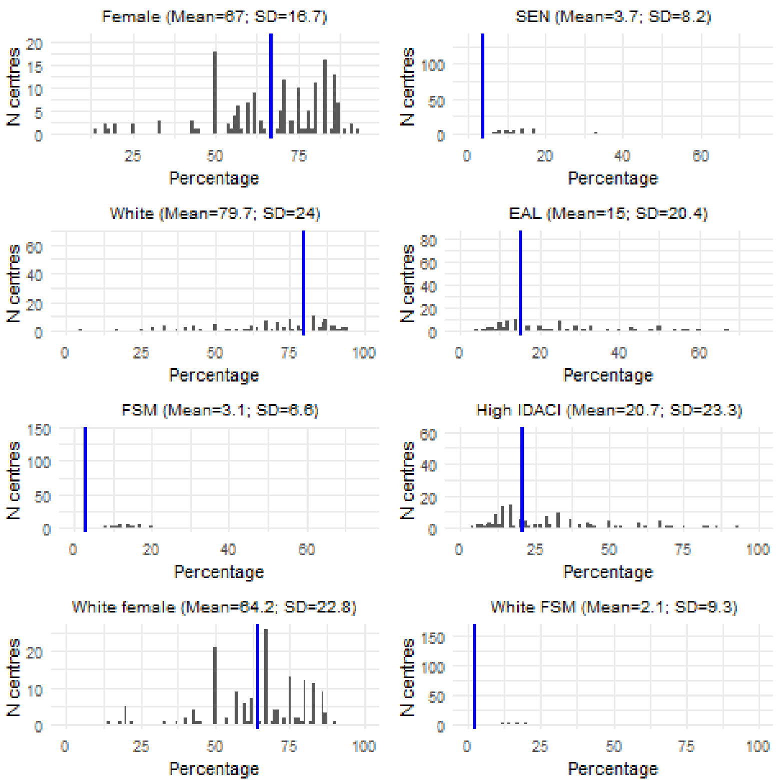


Figure F.14. Correlation between demographic/socio-economic factors and accuracy of predictions – French

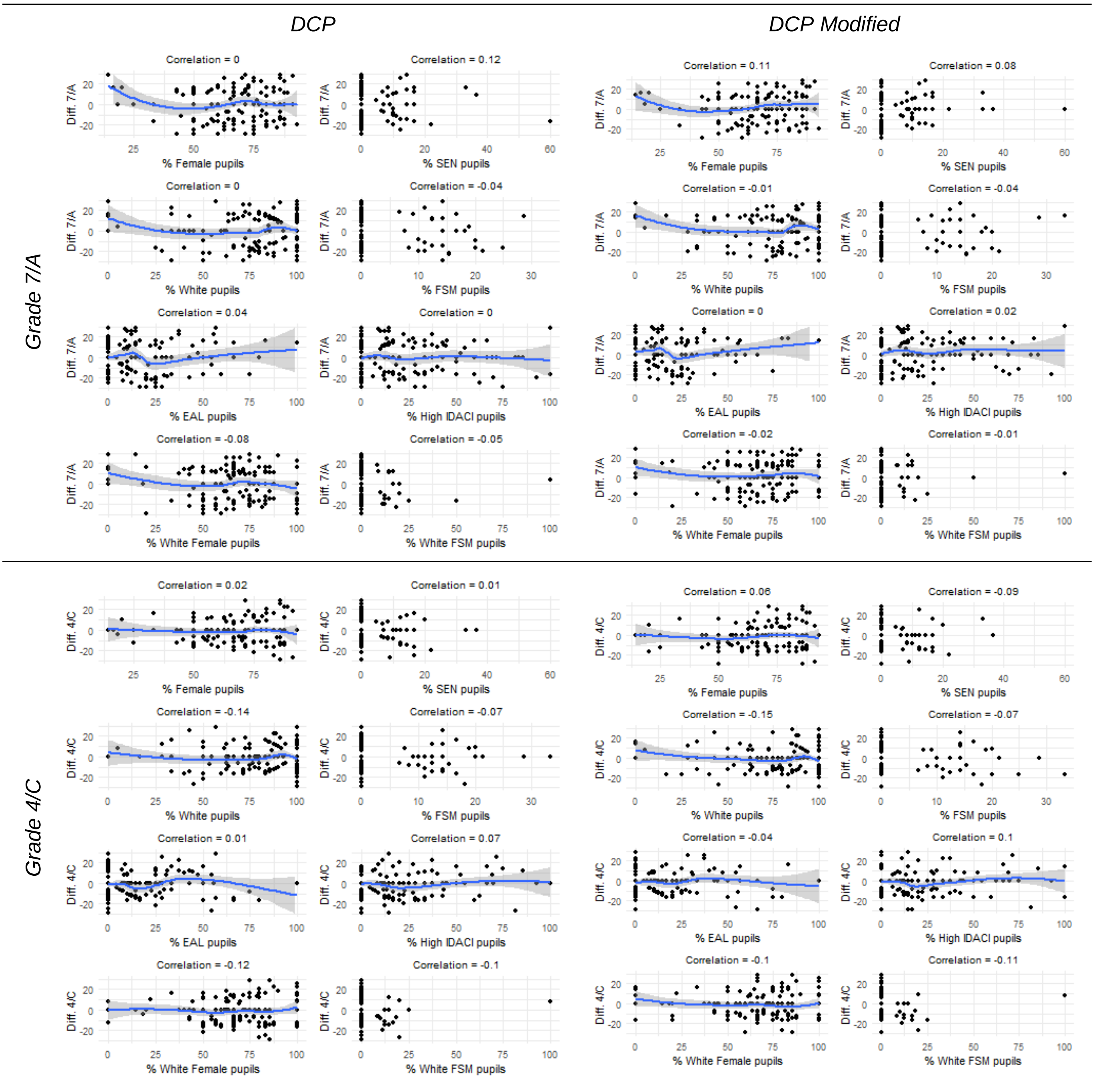


Table F.9. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – French

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	-7.47	-3.07	1.22	-12.15
	(17.37)	(13.68)	(15.23)	(12.06)
Perc.Female	0.12	0.13 *	0.13 *	0.15 *
	(0.15)	(0.12)	(0.13)	(0.10)
Perc.SEN	0.26	0.06	0.16	-0.12
	(0.19)	(0.15)	(0.17)	(0.13)
Perc.WHITE	0.03	-0.10	0.05	-0.11
	(0.11)	(0.08)	(0.09)	(0.07)
Perc.FSM	-0.06	-0.34	-0.16	-0.36
	(0.30)	(0.24)	(0.27)	(0.21)
Perc.EAL	0.06	-0.12	0.02	-0.12
	(0.10)	(0.08)	(0.09)	(0.07)
IDACI.high	-0.00	0.05	-0.01	0.07
	(0.10)	(0.08)	(0.08)	(0.07)
IDACI.med	-0.05	0.04	-0.13 *	-0.02
	(0.07)	(0.05)	(0.06)	(0.05)
WHITE_Female	-0.09	-0.11 *	-0.09	-0.12 **
	(0.11)	(0.09)	(0.10)	(0.08)
WHITE_FSM	-0.15	-0.27	0.01	-0.21
	(0.20)	(0.16)	(0.18)	(0.14)
mean.GCSE	0.17	-0.02	0.10	0.14
	(0.25)	(0.20)	(0.22)	(0.18)
Ncands_2019	-0.50	0.40	-0.78	0.21
	(0.48)	(0.38)	(0.42)	(0.33)
N	180	180	180	180
R2	0.05	0.10	0.09	0.14

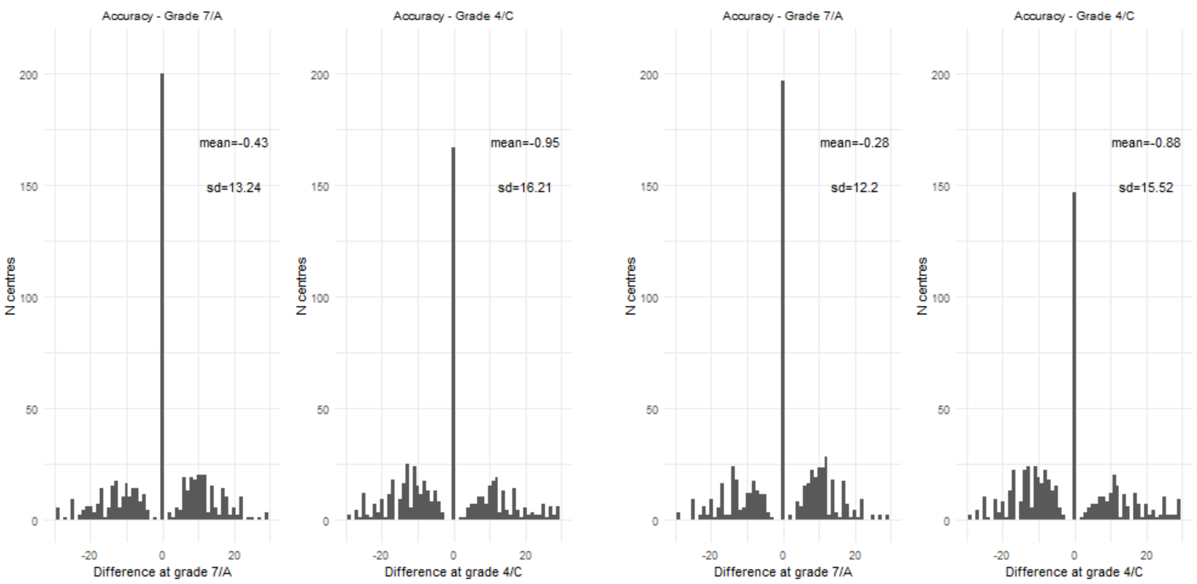
***** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.**

A level Religious studies

Figure F.15. Univariate distribution of key variables – Religious studies

a. Distribution of accuracy at centre level

DCP *DCP Modified*



b. Distribution of demographic and socio-economic centre composition

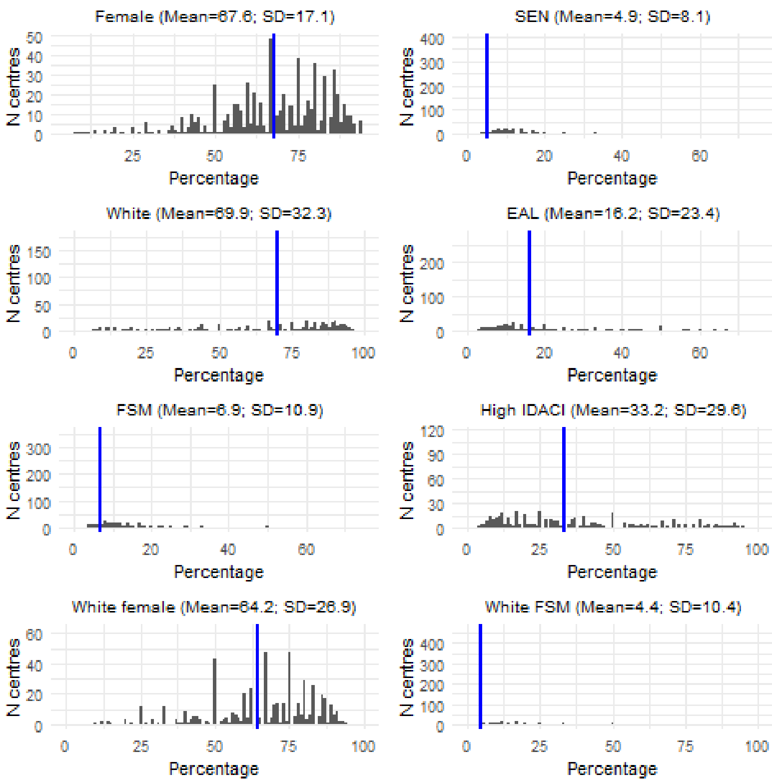


Figure F.16. Correlation between demographic/socio-economic factors and accuracy of predictions – Religious studies

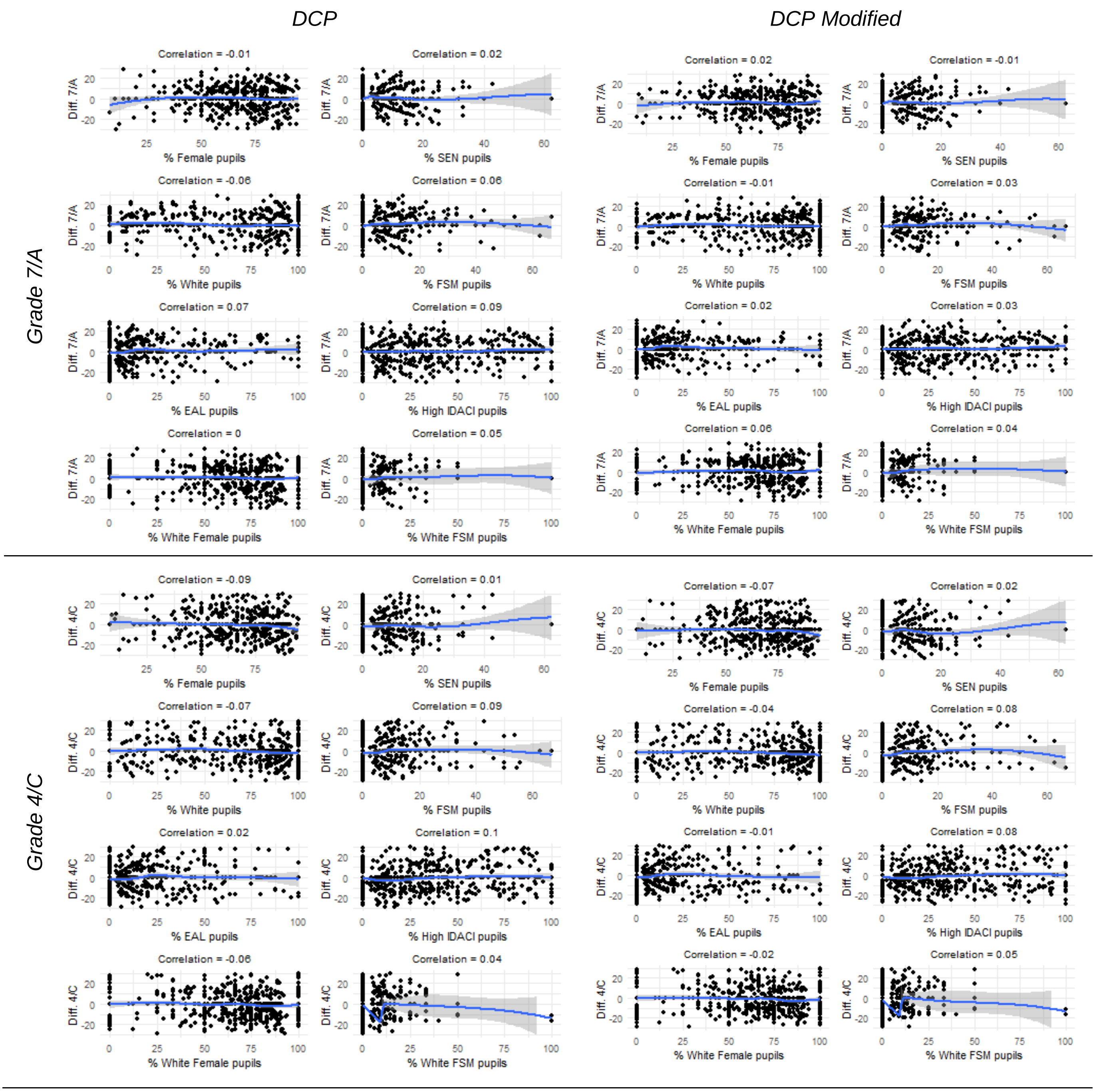


Table F.10. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Religious studies

	DCP		DCP Modified	
	Acc. 7/A+	**Acc. 4/C+**	**Acc. 7/A+**	**Acc. 4/C+**
(Intercept)	4.50	11.88	7.89	18.25 **
	(5.64)	(6.84)	(5.21)	(6.52)
Perc.Female	-0.02	-0.09	-0.01	-0.09
	(0.04)	(0.05)	(0.04)	(0.05)
Perc.SEN	0.02	-0.04	-0.03	-0.02
	(0.07)	(0.08)	(0.06)	(0.08)
Perc.WHITE	0.01	-0.03	0.01	-0.02
	(0.03)	(0.04)	(0.03)	(0.03)
Perc.FSM	-0.00	0.11	0.02	0.13
	(0.07)	(0.08)	(0.06)	(0.08)
Perc.EAL	0.02	-0.10 *	0.01	-0.11 **
	(0.04)	(0.04)	(0.03)	(0.04)
IDACI.high	0.02	0.04	0.01	0.05
	(0.03)	(0.04)	(0.03)	(0.03)
IDACI.med	-0.04	-0.03	-0.02	-0.01
	(0.03)	(0.04)	(0.03)	(0.04)
WHITE_Female	0.02	0.01	0.03	0.02
	(0.03)	(0.03)	(0.03)	(0.03)
WHITE_FSM	0.04	-0.02	0.02	-0.03

	(0.06)	(0.08)		(0.06)	(0.07)
mean.GCSE	-0.10	-0.25 *		-0.16	-0.35 **
	(0.10)	(0.12)		(0.09)	(0.11)
Ncands_2019	0.09	0.14		0.07	0.10
	(0.08)	(0.10)		(0.07)	(0.09)
N	630	630		630	630
R2	0.02	0.04		0.01	0.04

***** p < 0.001; ** p < 0.01; * p < 0.05.**

Annex G: Year-on-year variability

GCSE English language

Figure G.1. Year-on-year variability in outcomes – English language

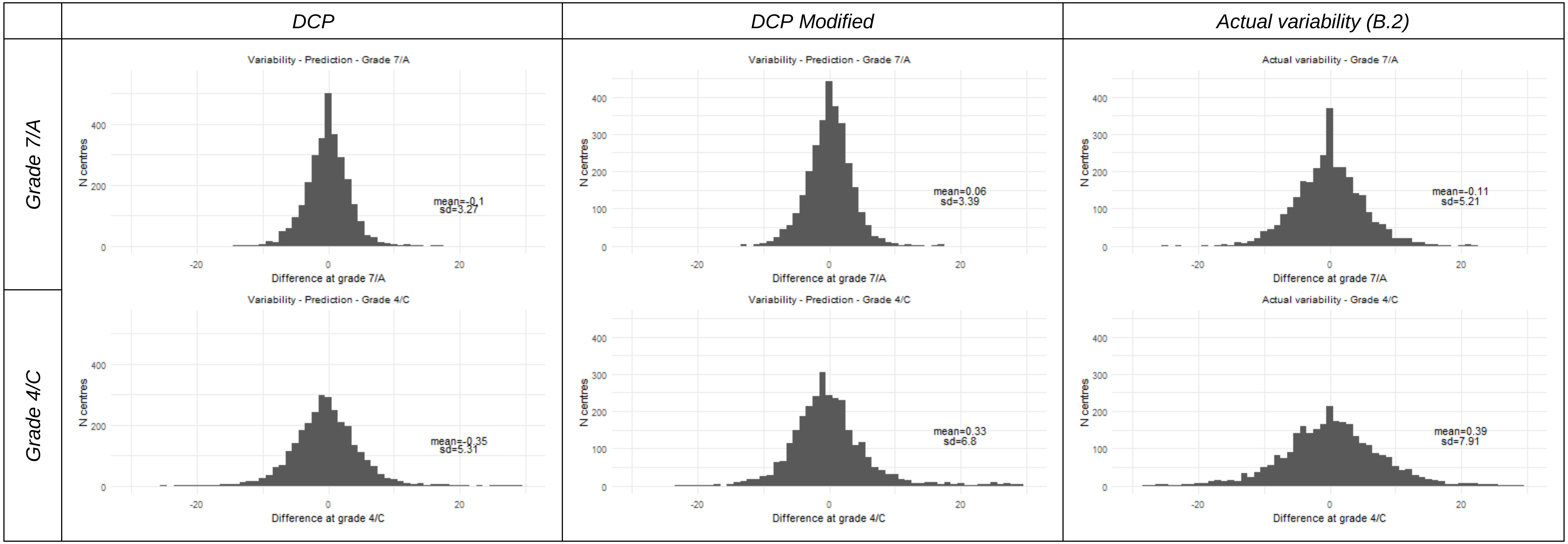


Figure G.2. Year-on-year variability in demographic and socio-economic centre composition – English language

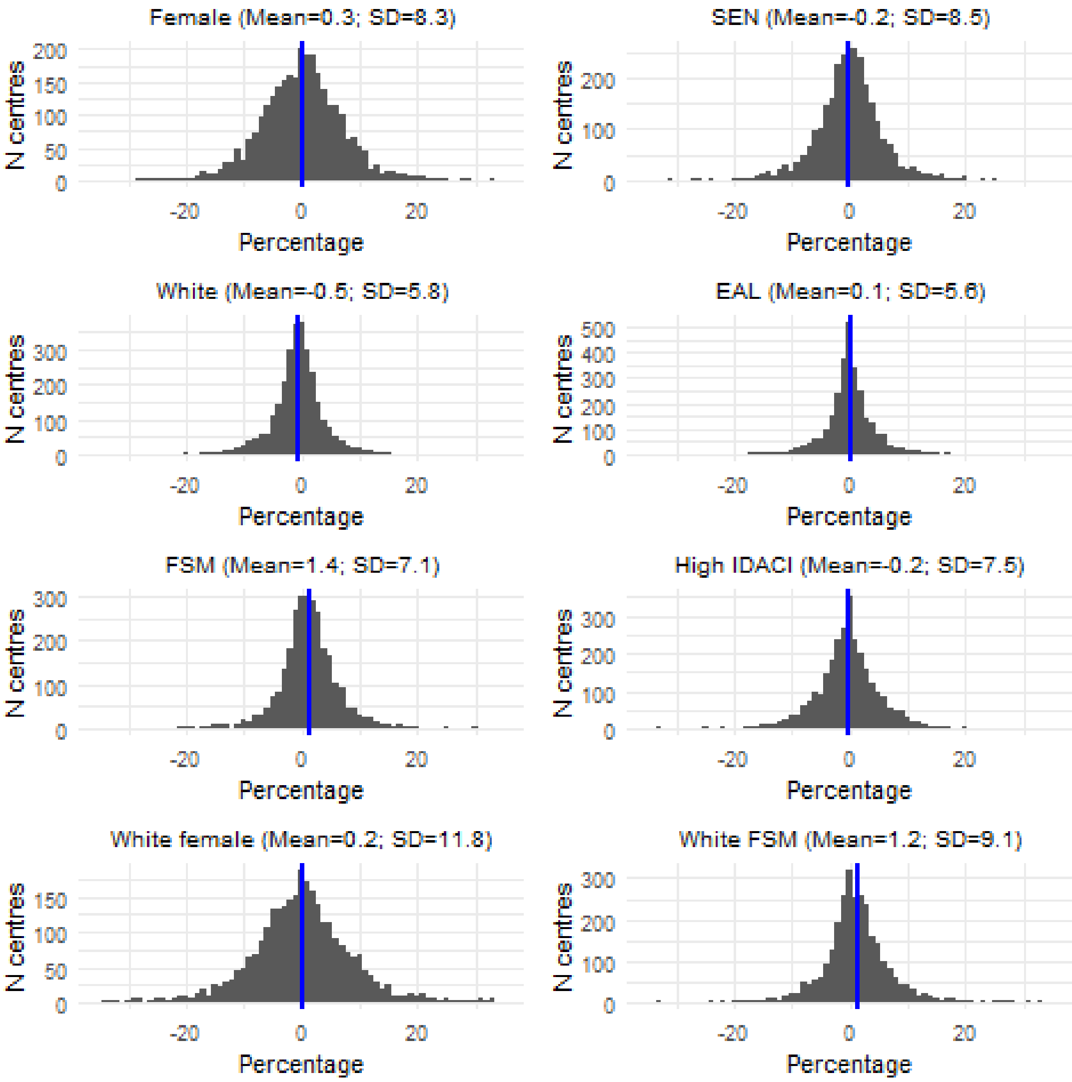
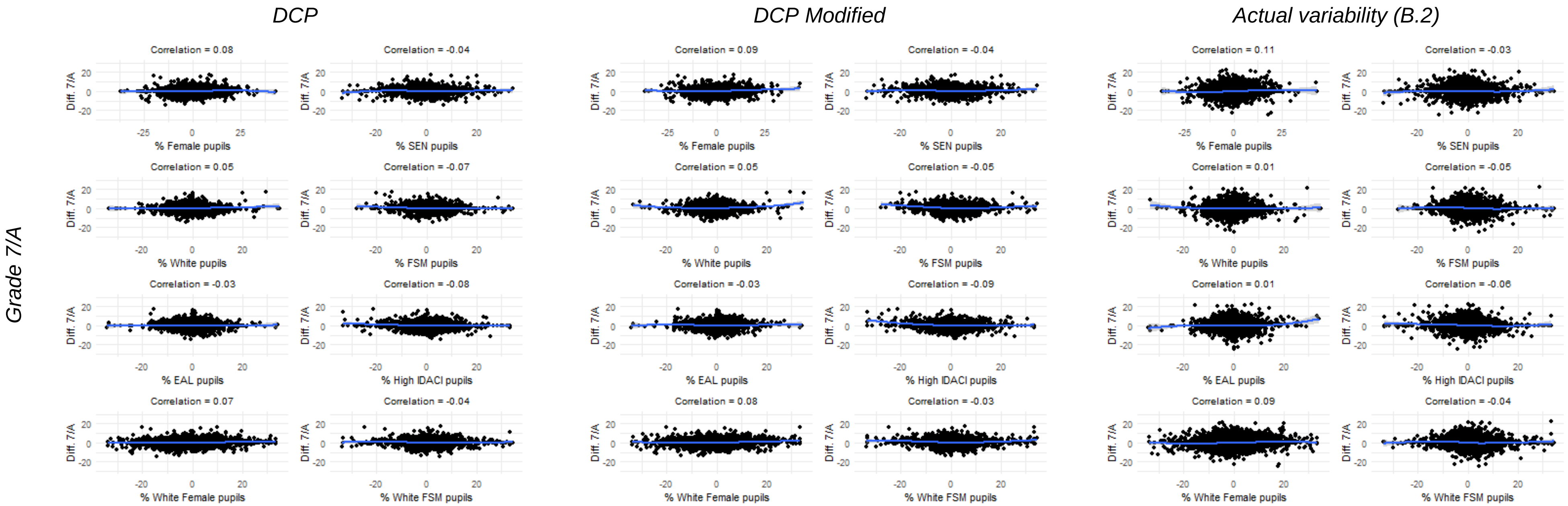


Figure G.3. Correlation between demographic/socio-economic factors and accuracy of predictions – English language



Grade 4/C

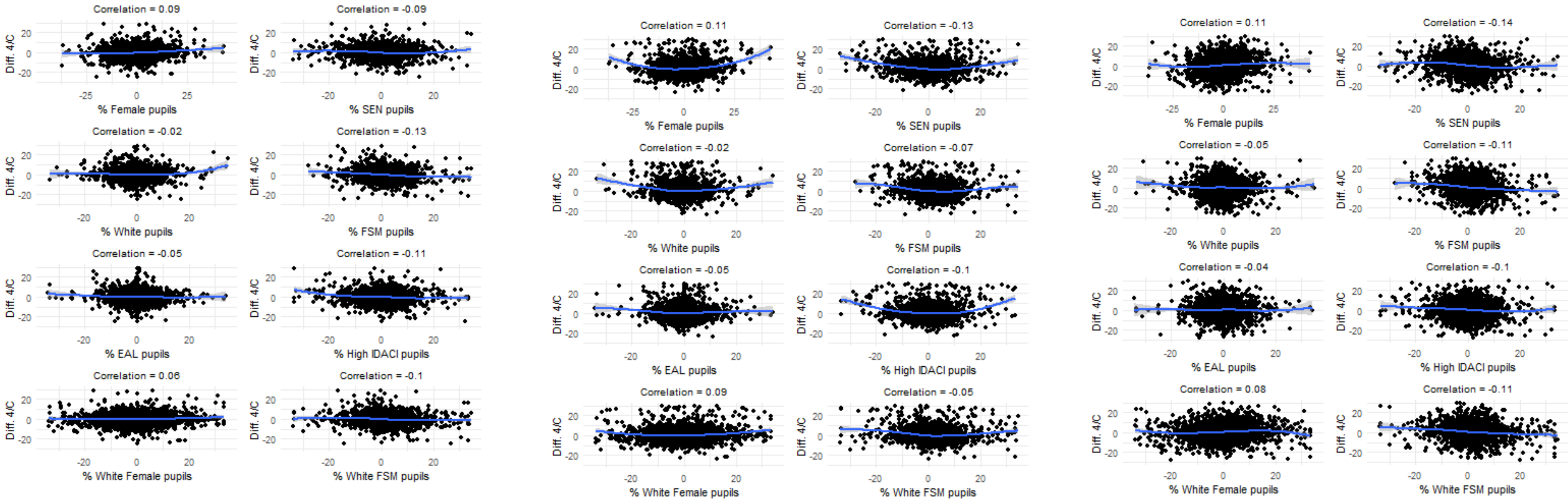

Correlation = 0.09
Correlation = -0.09
Correlation = -0.02
Correlation = -0.13
Correlation = -0.05
Correlation = -0.11
Correlation = 0.06
Correlation = -0.1
Correlation = 0.11
Correlation = -0.13
Correlation = -0.02
Correlation = -0.07
Correlation = -0.05
Correlation = -0.1
Correlation = 0.09
Correlation = -0.05
Correlation = 0.11
Correlation = -0.14
Correlation = -0.05
Correlation = -0.11
Correlation = -0.04
Correlation = -0.1
Correlation = 0.08
Correlation = -0.11
Diff. 4/C
% Female pupils
% SEN pupils
% White pupils
% FSM pupils
% EAL pupils
% High IDACI pupils
% White Female pupils
% White FSM pupils

Table G.1. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – English language

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	-0.30 *	-0.35	0.28 *	3.24 ***	-0.38	0.45
	(0.13)	(0.19)	(0.13)	(0.25)	(0.21)	(0.31)
Perc.Female	0.02	0.06 ***	0.02 *	0.08 ***	0.05 **	0.12 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Perc.SEN	0.00	-0.01	0.00	-0.07 ***	0.00	-0.09 ***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Perc.WHITE	0.02 *	-0.03	0.02 *	-0.04 *	0.01	-0.08 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Perc.FSM	0.00	-0.02	0.02	-0.00	0.02	-0.01
	(0.01)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)
Perc.EAL	0.00	-0.02	0.00	-0.04	0.03	-0.04
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
IDACI.high	-0.03 **	-0.03 *	-0.04 ***	-0.03	-0.07 ***	-0.10 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
IDACI.med	-0.02	0.03	-0.01	0.06 **	-0.05 **	-0.04
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
WHITE_Female	0.01	-0.01	0.01	0.01	0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
WHITE_FSM	0.01	0.00	0.01	0.01	-0.00	-0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
mean.ks2	8.98 ***	17.39 ***	9.60 ***	17.67 ***	9.05 ***	16.33 ***
	(0.43)	(0.67)	(0.45)	(0.87)	(0.72)	(1.06)
Ncands_2019	0.00	-0.00	-0.00	-0.02 ***	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	2928	2928	2928	2928	2928	2928
R2	0.15	0.22	0.16	0.21	0.07	0.13

***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE Mathematics

Figure G.4. Year-on-year variability in outcomes – Maths

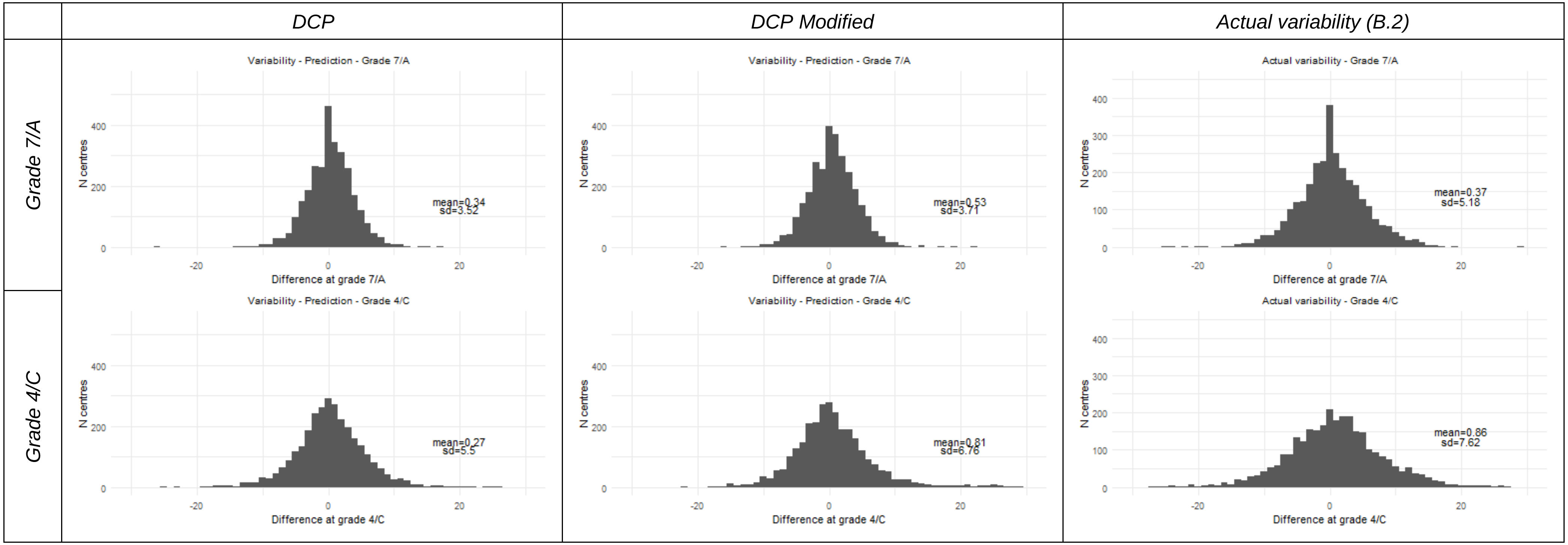


Figure G.5. Year-on-year variability in demographic and socio-economic centre composition – Maths

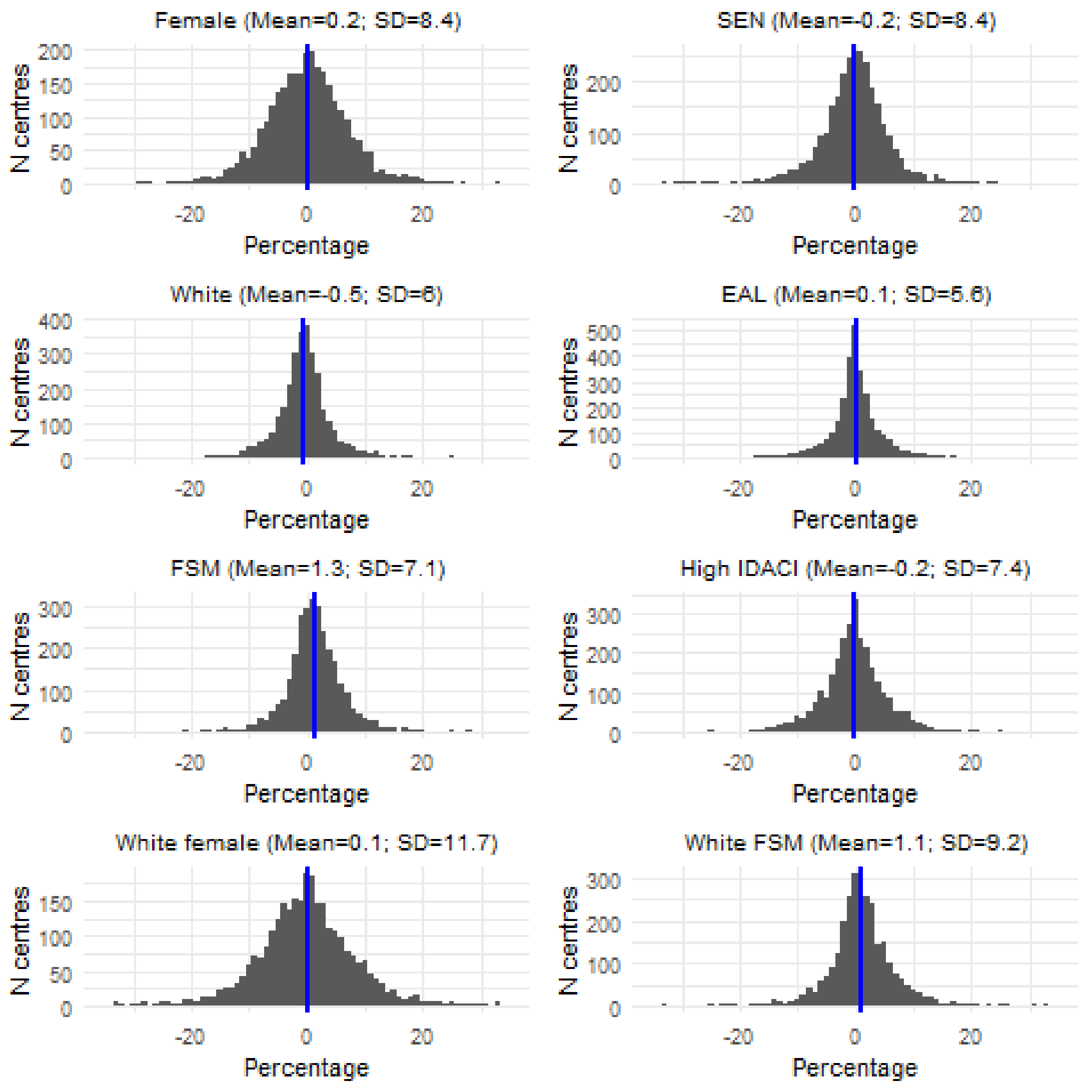
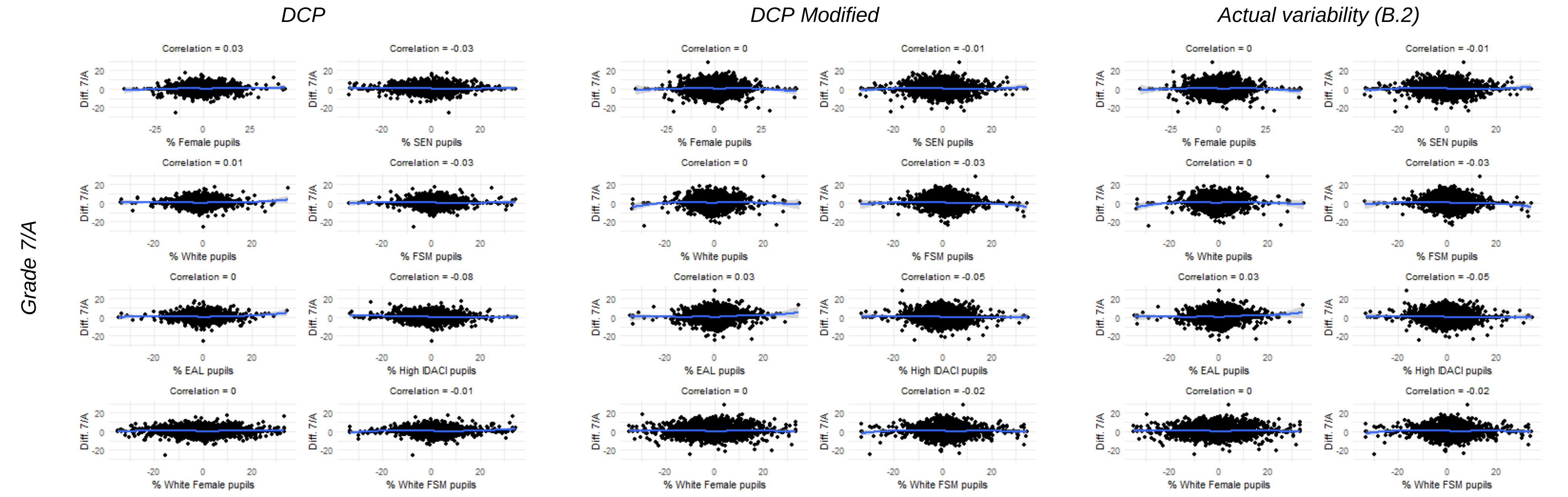


Figure G.6. Correlation between demographic/socio-economic factors and accuracy of predictions – Maths



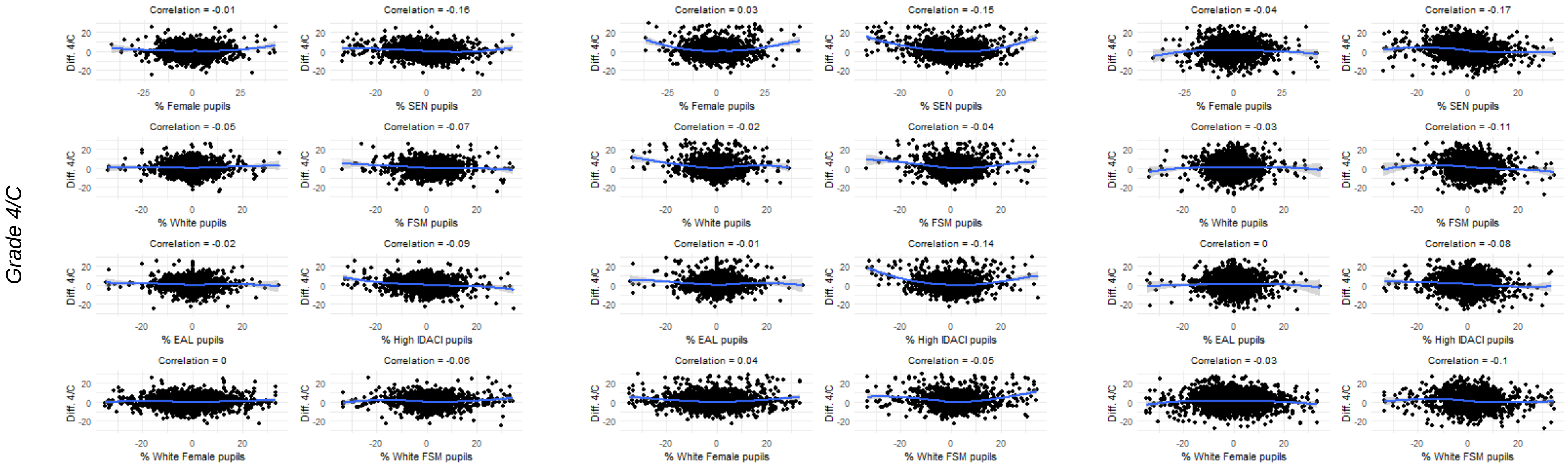

Grade 4/C
Correlation = -0.01
% Female pupils
Correlation = -0.16
% SEN pupils
Correlation = 0.03
% Female pupils
Correlation = -0.15
% SEN pupils
Correlation = -0.04
% Female pupils
Correlation = -0.17
% SEN pupils
Correlation = -0.05
% White pupils
Correlation = -0.07
% FSM pupils
Correlation = -0.02
% White pupils
Correlation = -0.04
% FSM pupils
Correlation = -0.03
% White pupils
Correlation = -0.11
% FSM pupils
Correlation = -0.02
% EAL pupils
Correlation = -0.09
% High IDACI pupils
Correlation = -0.01
% EAL pupils
Correlation = -0.14
% High IDACI pupils
Correlation = 0
% EAL pupils
Correlation = -0.08
% High IDACI pupils
Correlation = 0
% White Female pupils
Correlation = -0.06
% White FSM pupils
Correlation = 0.04
% White Female pupils
Correlation = -0.05
% White FSM pupils
Correlation = -0.03
% White Female pupils
Correlation = -0.1
% White FSM pupils
Diff. 4/C

Table G.2. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Maths

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	0.08	0.21	0.81 ***	3.23 ***	0.13	1.24 ***
	(0.13)	(0.19)	(0.14)	(0.24)	(0.20)	(0.29)
Perc.Female	0.02 *	-0.00	0.01	0.01	-0.00	-0.02
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Perc.SEN	0.02 **	-0.04 ***	0.01	-0.06 ***	0.02 *	-0.10 ***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Perc.WHITE	0.01	-0.04 *	0.02	-0.04 *	-0.00	-0.04
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
Perc.FSM	0.02	0.03	0.03	0.08 **	0.00	-0.01
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
Perc.EAL	0.01	-0.00	0.02	-0.00	0.04 *	0.00
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
IDACI.high	-0.03 **	-0.07 ***	-0.03 *	-0.12 ***	-0.07 ***	-0.12 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
IDACI.med	-0.01	-0.04 *	-0.01	-0.03	-0.06 ***	-0.10 ***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
WHITE_Female	-0.01	-0.00	0.00	0.01	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
WHITE_FSM	0.01	0.01	0.01	-0.03	0.01	-0.01
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)
mean.ks2	10.89 ***	18.99 ***	11.64 ***	20.04 ***	9.03 ***	16.18 ***
	(0.46)	(0.69)	(0.49)	(0.86)	(0.72)	(1.03)
Ncands_2019	0.00 *	0.00	-0.00 *	-0.01 ***	0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	2939	2939	2939	2939	2939	2939
R2	0.17	0.24	0.18	0.23	0.06	0.12

***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE Music

Figure G.7. Year-on-year variability in outcomes – Music

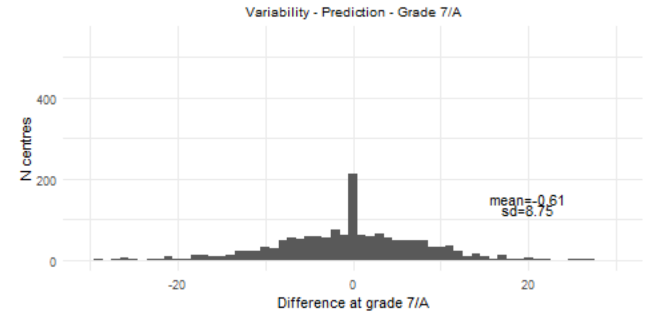
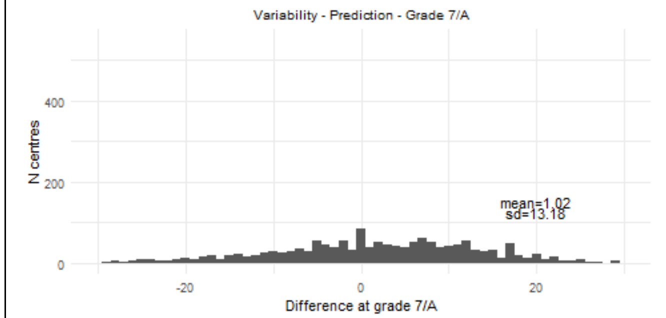
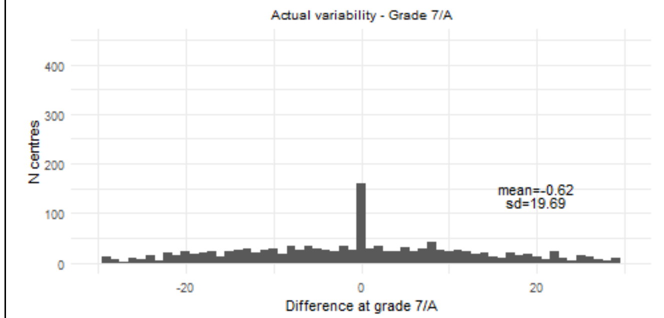
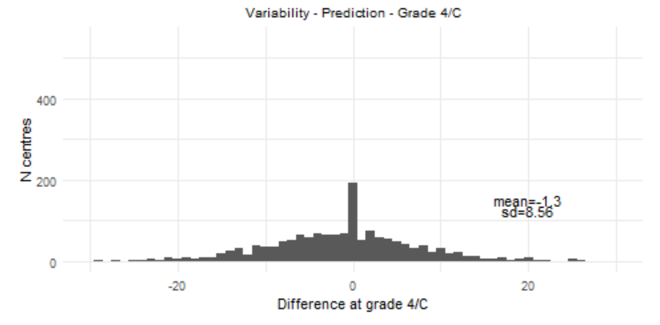
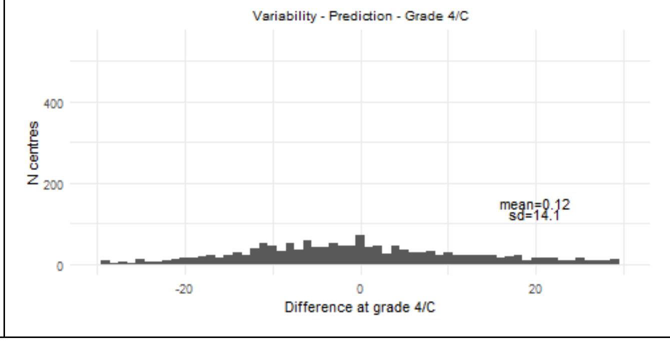
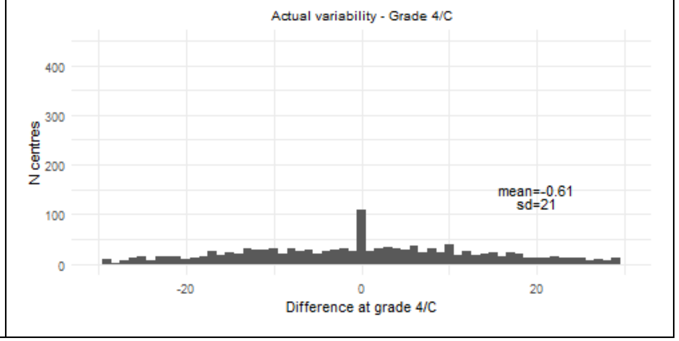
	DCP	*DCP Modified*	*Actual variability (B.2)*
Grade 7/A			
Grade 4/C			

Figure G.8. Year-on-year variability in demographic and socio-economic centre composition – Music

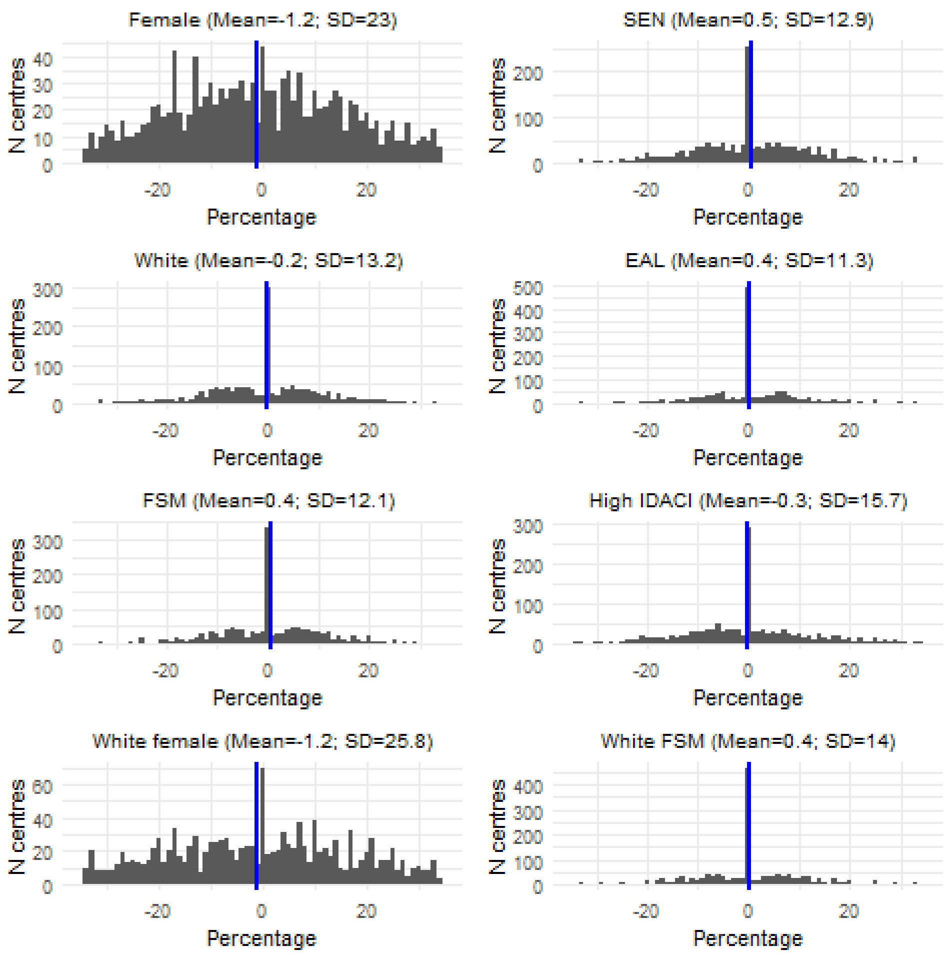
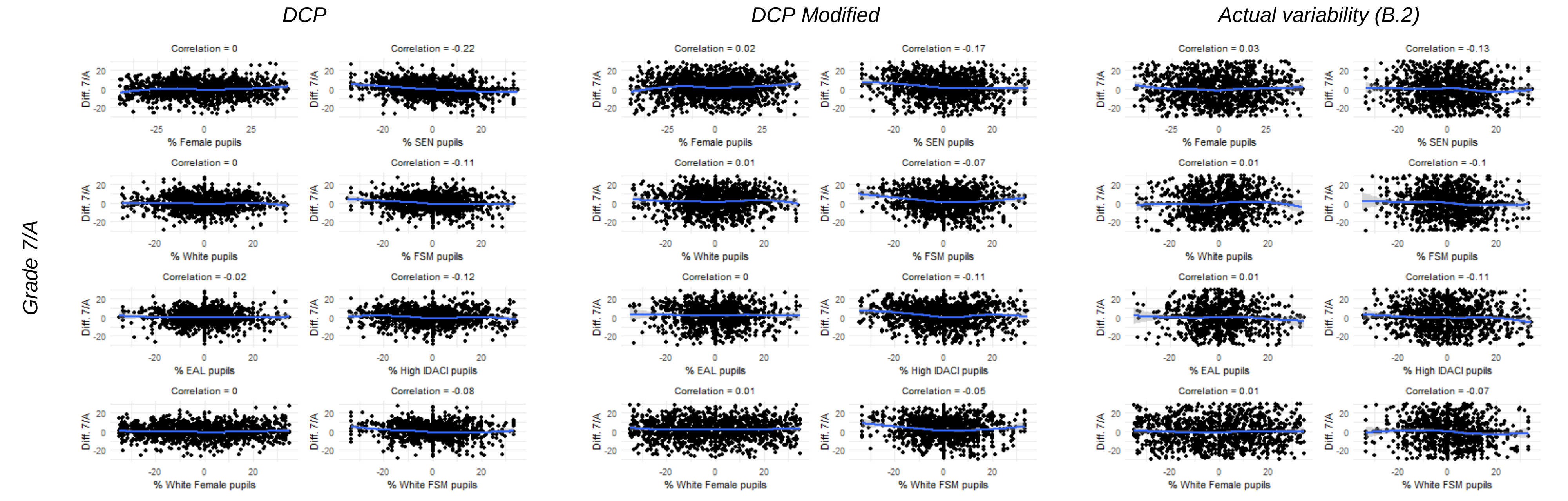


Figure G.9. Correlation between demographic/socio-economic factors and accuracy of predictions – Music



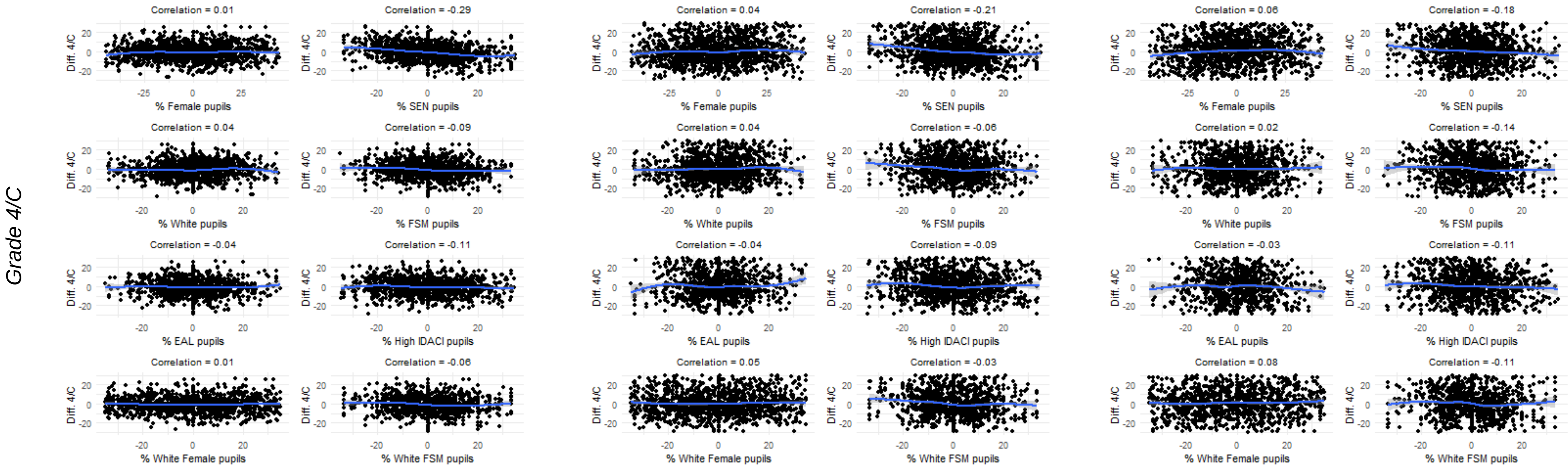

Grade 4/C
Correlation = 0.01
Diff. 4/C
% Female pupils
Correlation = -0.29
Diff. 4/C
% SEN pupils
Correlation = 0.04
Diff. 4/C
% Female pupils
Correlation = -0.21
Diff. 4/C
% SEN pupils
Correlation = 0.06
Diff. 4/C
% Female pupils
Correlation = -0.18
Diff. 4/C
% SEN pupils
Correlation = 0.04
Diff. 4/C
% White pupils
Correlation = -0.09
Diff. 4/C
% FSM pupils
Correlation = 0.04
Diff. 4/C
% White pupils
Correlation = -0.06
Diff. 4/C
% FSM pupils
Correlation = 0.02
Diff. 4/C
% White pupils
Correlation = -0.14
Diff. 4/C
% FSM pupils
Correlation = -0.04
Diff. 4/C
% EAL pupils
Correlation = -0.11
Diff. 4/C
% High IDACI pupils
Correlation = -0.04
Diff. 4/C
% EAL pupils
Correlation = -0.09
Diff. 4/C
% High IDACI pupils
Correlation = -0.03
Diff. 4/C
% EAL pupils
Correlation = -0.11
Diff. 4/C
% High IDACI pupils
Correlation = 0.01
Diff. 4/C
% White Female pupils
Correlation = -0.06
Diff. 4/C
% White FSM pupils
Correlation = 0.05
Diff. 4/C
% White Female pupils
Correlation = -0.03
Diff. 4/C
% White FSM pupils
Correlation = 0.08
Diff. 4/C
% White Female pupils
Correlation = -0.11
Diff. 4/C
% White FSM pupils

Table G.3. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Music

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	-0.07	-1.27 ***	3.38 ***	2.46 ***	0.83	1.19
	(0.40)	(0.33)	(0.69)	(0.72)	(1.10)	(1.16)
Perc.Female	0.00	-0.02	0.01	-0.00	0.06	-0.04
	(0.02)	(0.01)	(0.03)	(0.03)	(0.04)	(0.05)
Perc.SEN	-0.02	-0.05 ***	-0.03	-0.09 ***	-0.06	-0.14 ***
	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)	(0.04)
Perc.WHITE	0.00	0.03 **	0.01	0.04	0.00	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.04)
Perc.FSM	-0.03	-0.02	-0.02	-0.06	-0.15 *	-0.19 **
	(0.02)	(0.02)	(0.04)	(0.04)	(0.06)	(0.07)
Perc.EAL	-0.01	-0.02	0.01	-0.04	0.03	-0.06
	(0.02)	(0.01)	(0.03)	(0.03)	(0.04)	(0.05)
IDACI.high	-0.03 *	-0.01	-0.07 ***	-0.05 *	-0.14 ***	-0.13 ***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.04)
IDACI.med	-0.02	-0.02 *	-0.06 ***	-0.05 *	-0.11 ***	-0.09 **
	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)
WHITE_Female	0.00	0.02	0.00	0.03	-0.03	0.11 *
	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)	(0.04)
WHITE_FSM	0.01	0.03	0.02	0.07 *	0.07	0.04
	(0.02)	(0.02)	(0.03)	(0.04)	(0.05)	(0.06)
mean.ks2	20.91 ***	24.15 ***	22.94 ***	25.03 ***	21.90 ***	22.62 ***
	(0.70)	(0.58)	(1.21)	(1.28)	(1.95)	(2.05)
Ncands_2019	-0.03	-0.00	-0.15 ***	-0.15 ***	-0.09	-0.11
	(0.02)	(0.02)	(0.04)	(0.04)	(0.06)	(0.07)
N	1496	1496	1496	1496	1496	1496
R2	0.42	0.60	0.24	0.27	0.13	0.15

***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE History

Figure G.10. Year-on-year variability in outcomes – History

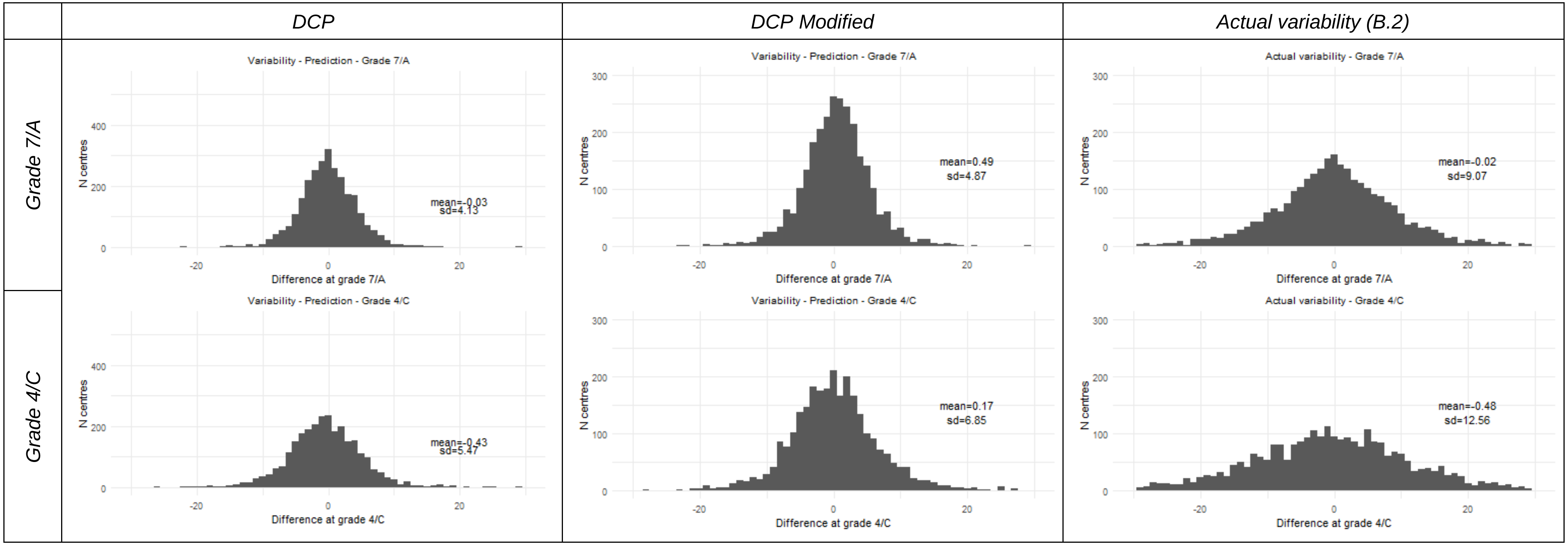


Figure G.11. Year-on-year variability in demographic and socio-economic centre composition – History

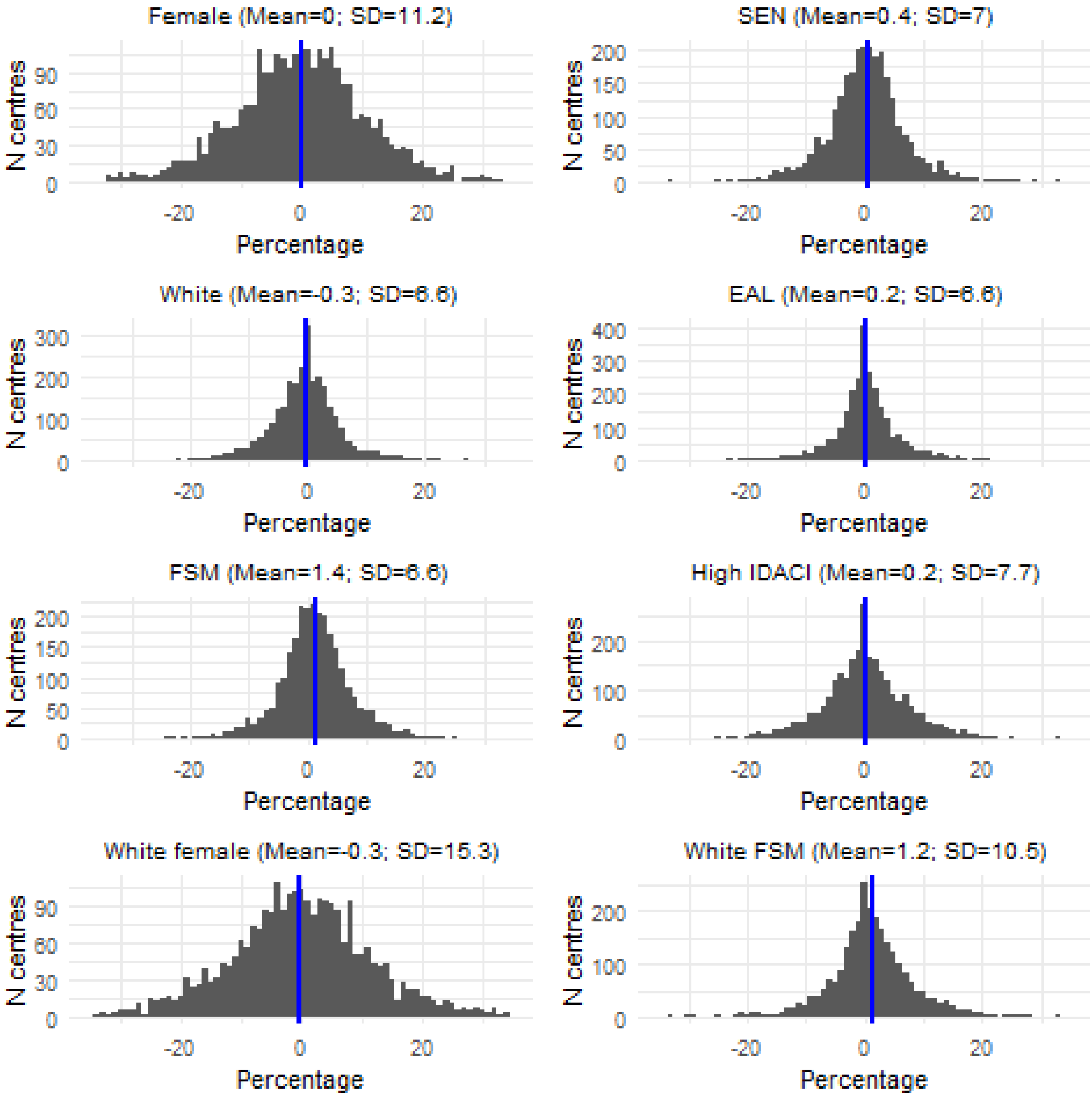
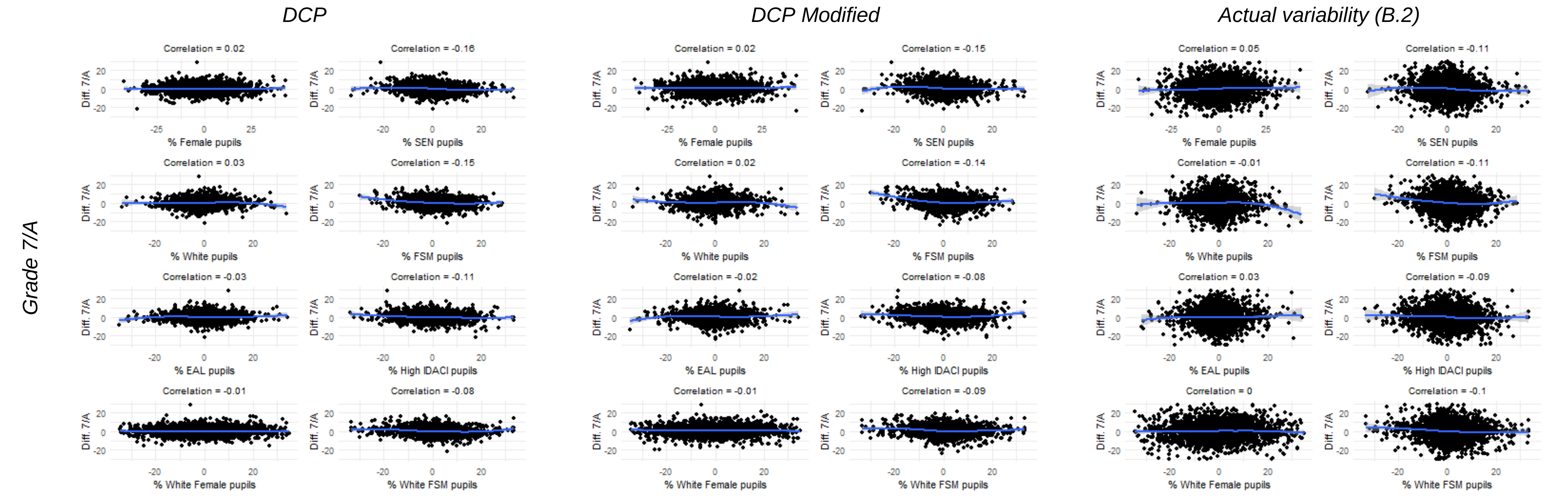


Figure G.12. Correlation between demographic/socio-economic factors and accuracy of predictions – History



Grade 4/C

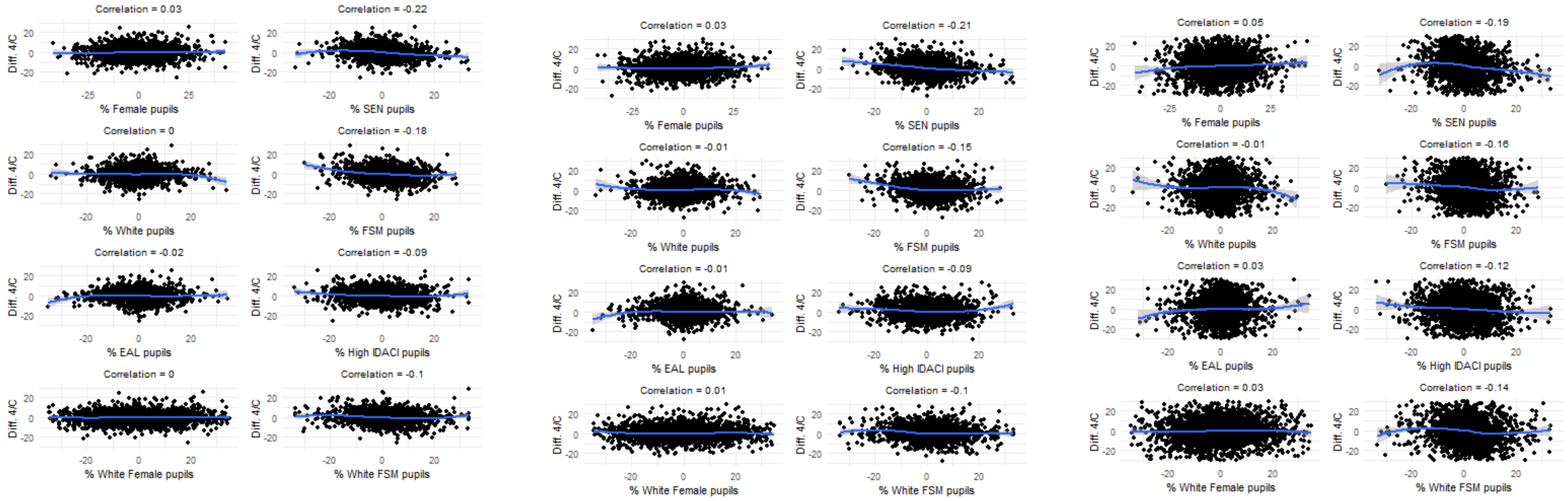

Correlation = 0.03
Diff. 4/C
% Female pupils
Correlation = -0.22
% SEN pupils
Correlation = 0
% White pupils
Correlation = -0.18
% FSM pupils
Correlation = -0.02
% EAL pupils
Correlation = -0.09
% High IDACI pupils
Correlation = 0
% White Female pupils
Correlation = -0.1
% White FSM pupils
Correlation = 0.03
% Female pupils
Correlation = -0.21
% SEN pupils
Correlation = -0.01
% White pupils
Correlation = -0.15
% FSM pupils
Correlation = -0.01
% EAL pupils
Correlation = -0.09
% High IDACI pupils
Correlation = 0.01
% White Female pupils
Correlation = -0.1
% White FSM pupils
Correlation = 0.05
% Female pupils
Correlation = -0.19
% SEN pupils
Correlation = -0.01
% White pupils
Correlation = -0.16
% FSM pupils
Correlation = 0.03
% EAL pupils
Correlation = -0.12
% High IDACI pupils
Correlation = 0.03
% White Female pupils
Correlation = -0.14
% White FSM pupils

Table G.4. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – History

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	0.34 **	-0.25	1.71 ***	1.69 ***	0.26	-0.48
	(0.12)	(0.14)	(0.16)	(0.22)	(0.38)	(0.51)
Perc.Female	0.00	-0.00	0.00	-0.01	0.06 **	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Perc.SEN	0.01	-0.03 **	0.00	-0.05 ***	-0.02	-0.15 ***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Perc.WHITE	0.02 *	0.00	0.02	-0.02	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)
Perc.FSM	-0.01	-0.02	-0.00	-0.02	-0.00	-0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)
Perc.EAL	-0.01	-0.02 *	-0.01	-0.02	0.06 *	0.05
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)
IDACI.high	-0.04 ***	-0.03 **	-0.03 **	-0.06 ***	-0.13 ***	-0.20 ***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)
IDACI.med	-0.01	-0.00	-0.01	-0.02	-0.07 **	-0.09 **
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
WHITE_Female	-0.01	-0.00	-0.01	0.00	-0.03 *	0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
WHITE_FSM	0.01 *	0.01	0.00	0.00	-0.04	-0.08 **
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
mean.ks2	22.58 ***	31.56 ***	23.60 ***	32.57 ***	22.93 ***	34.25 ***
	(0.37)	(0.43)	(0.49)	(0.69)	(1.18)	(1.59)
Ncands_2019	-0.00 ***	-0.00	-0.02 ***	-0.02 ***	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
N	2712	2712	2712	2712	2712	2712
R2	0.61	0.69	0.49	0.49	0.16	0.21

***** p < 0.001; ** p < 0.01; * p < 0.05.**

GCSE Geography

Figure G.13. Year-on-year variability in outcomes – Geography

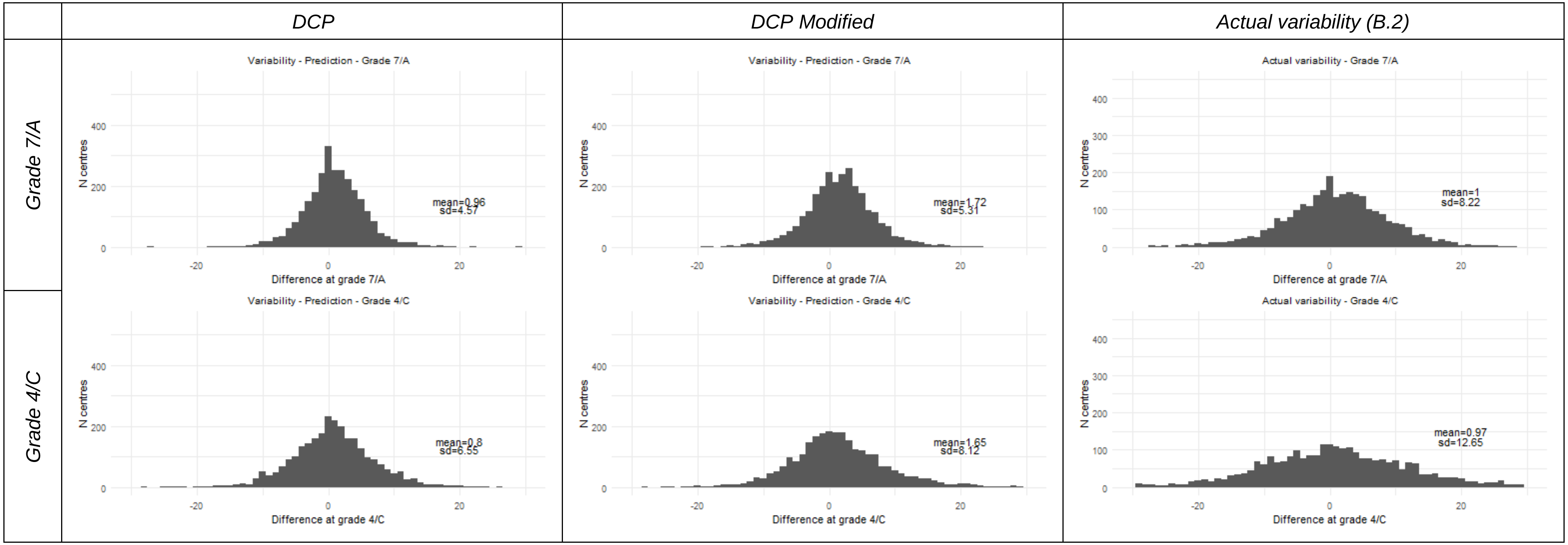


Figure G.14. Year-on-year variability in demographic and socio-economic centre composition – Geography

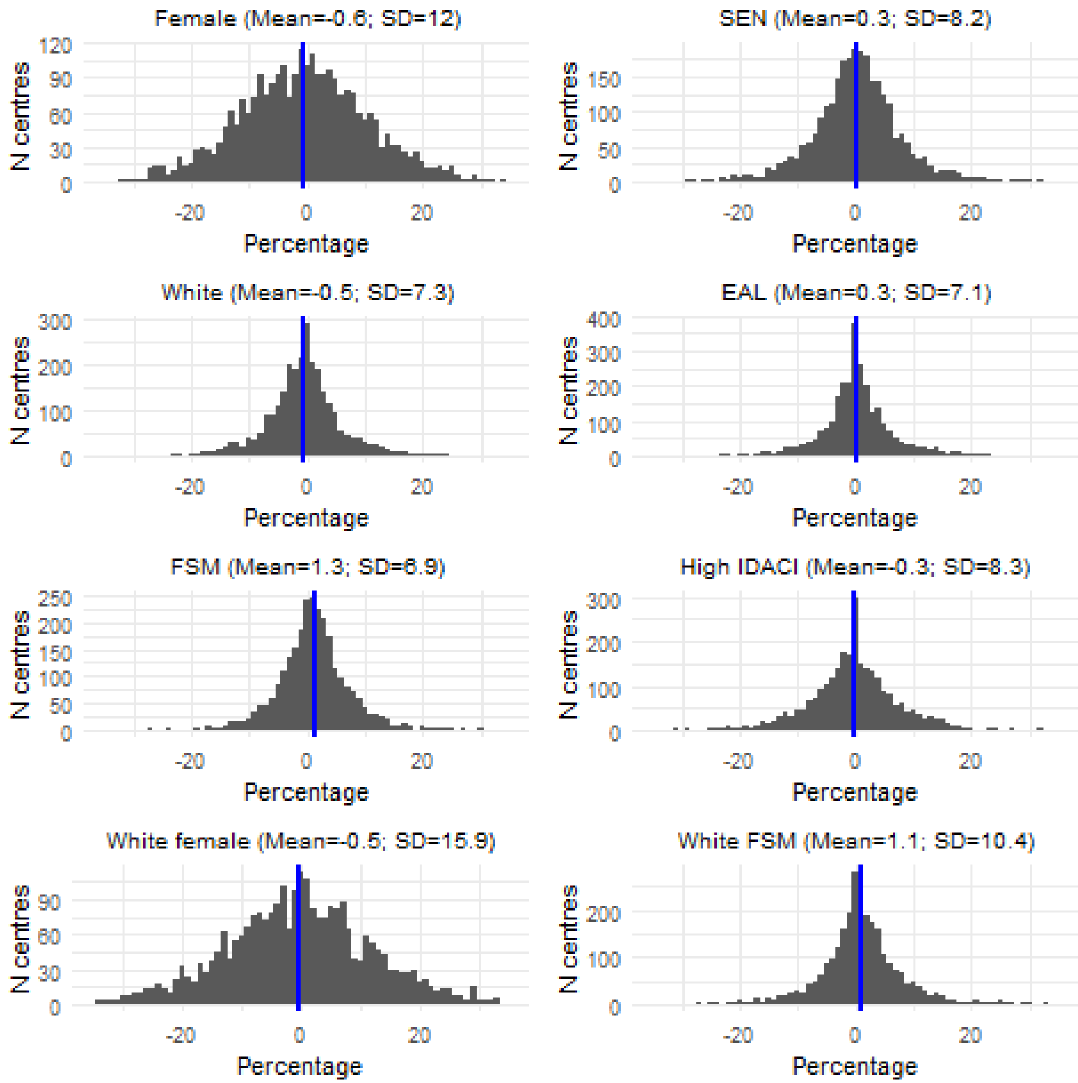
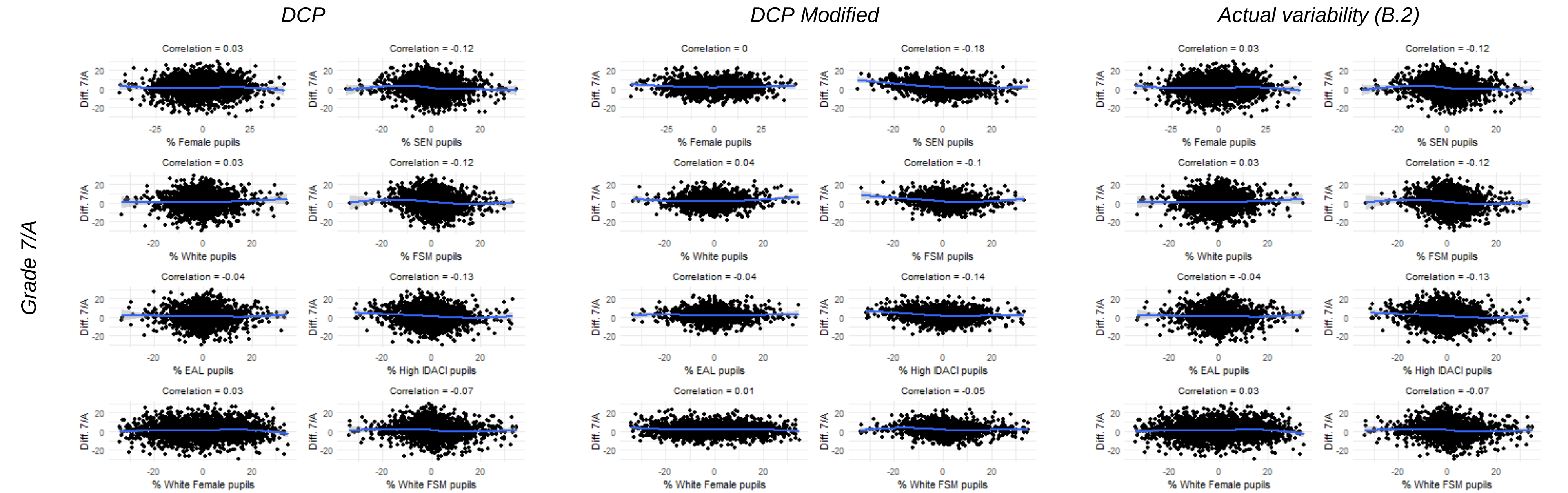


Figure G.15. Correlation between demographic/socio-economic factors and accuracy of predictions – Geography



Grade 4/C

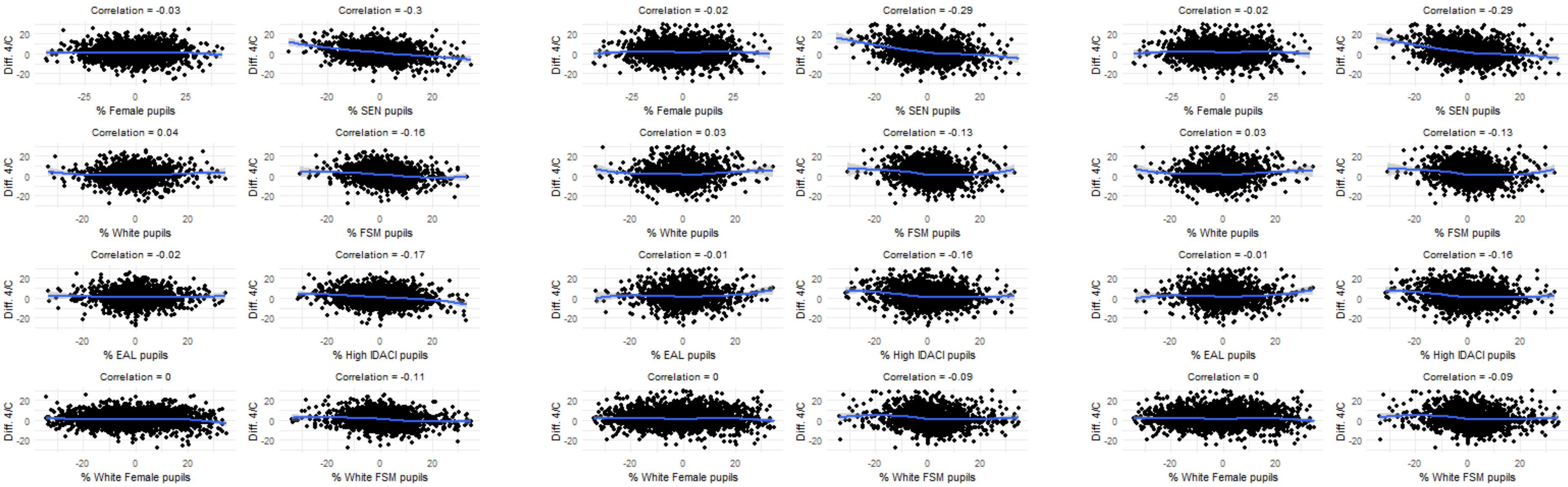

Correlation = -0.03
% Female pupils
Correlation = -0.3
% SEN pupils
Correlation = -0.02
% Female pupils
Correlation = -0.29
% SEN pupils
Correlation = -0.02
% Female pupils
Correlation = -0.29
% SEN pupils
Correlation = 0.04
% White pupils
Correlation = -0.16
% FSM pupils
Correlation = 0.03
% White pupils
Correlation = -0.13
% FSM pupils
Correlation = 0.03
% White pupils
Correlation = -0.13
% FSM pupils
Correlation = -0.02
% EAL pupils
Correlation = -0.17
% High IDACI pupils
Correlation = -0.01
% EAL pupils
Correlation = -0.16
% High IDACI pupils
Correlation = -0.01
% EAL pupils
Correlation = -0.16
% High IDACI pupils
Correlation = 0
% White Female pupils
Correlation = -0.11
% White FSM pupils
Correlation = 0
% White Female pupils
Correlation = -0.09
% White FSM pupils
Correlation = 0
% White Female pupils
Correlation = -0.09
% White FSM pupils
Diff. 4/C

Table G.5. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Geography

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	1.20 ***	0.98 ***	3.05 ***	3.83 ***	1.25 ***	1.75 ***
	(0.13)	(0.15)	(0.16)	(0.23)	(0.32)	(0.47)
Perc.Female	-0.01	-0.02 **	-0.01	-0.03 **	0.02	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Perc.SEN	-0.02 *	-0.09 ***	-0.01	-0.12 ***	-0.01	-0.19 ***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Perc.WHITE	0.00	0.01	0.00	0.02	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
Perc.FSM	-0.01	-0.02	0.00	0.00	-0.04	-0.05
	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.04)
Perc.EAL	-0.01	0.02	-0.01	0.03	-0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
IDACI.high	-0.00	-0.01	-0.02 *	-0.02	-0.10 ***	-0.18 ***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
IDACI.med	0.01	0.03 **	-0.00	0.03	-0.08 ***	-0.07 *
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
WHITE_Female	-0.00	-0.00	0.00	0.01	0.00	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
WHITE_FSM	0.00	-0.01	0.00	-0.01	-0.00	-0.05
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)
mean.ks2	21.36 ***	33.04 ***	23.18 ***	36.39 ***	21.53 ***	37.48 ***
	(0.37)	(0.43)	(0.46)	(0.65)	(0.92)	(1.32)
Ncands_2019	-0.00 *	-0.00	-0.02 ***	-0.03 ***	-0.00	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
N	2723	2723	2723	2723	2723	2723
R2	0.58	0.72	0.52	0.60	0.21	0.31

***** p < 0.001; ** p < 0.01; * p < 0.05.**

A level Biology

Figure G.16. Year-on-year variability in outcomes – Biology

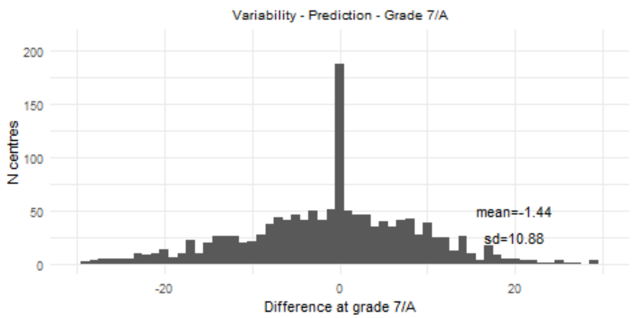
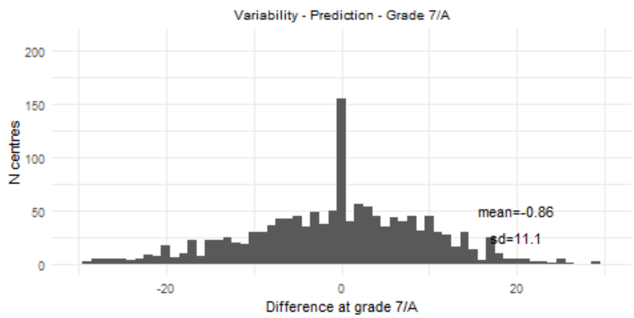
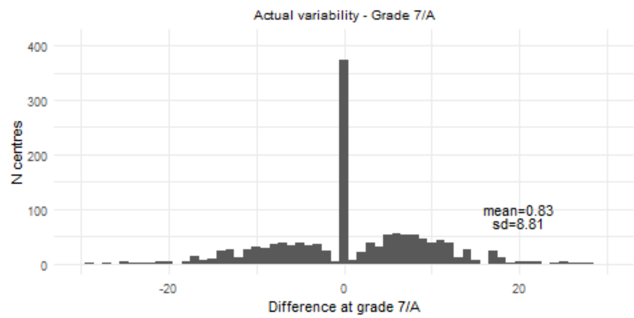
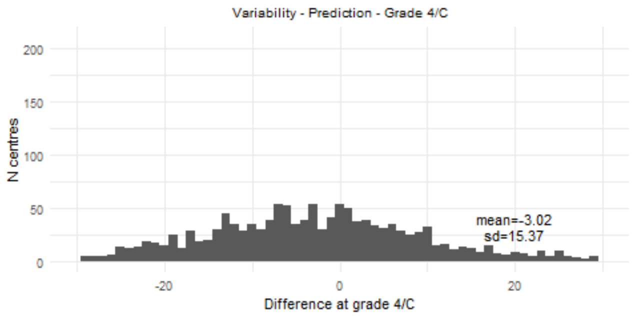
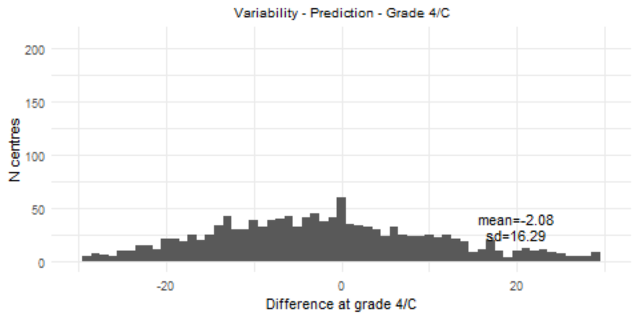
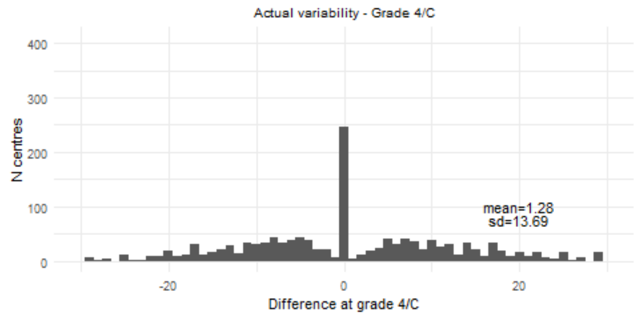
	DCP	*DCP Modified*	*Actual variability (B.2)*
Grade 7/A			
Grade 4/C			

Figure G.17. Year-on-year variability in demographic and socio-economic centre composition – Biology

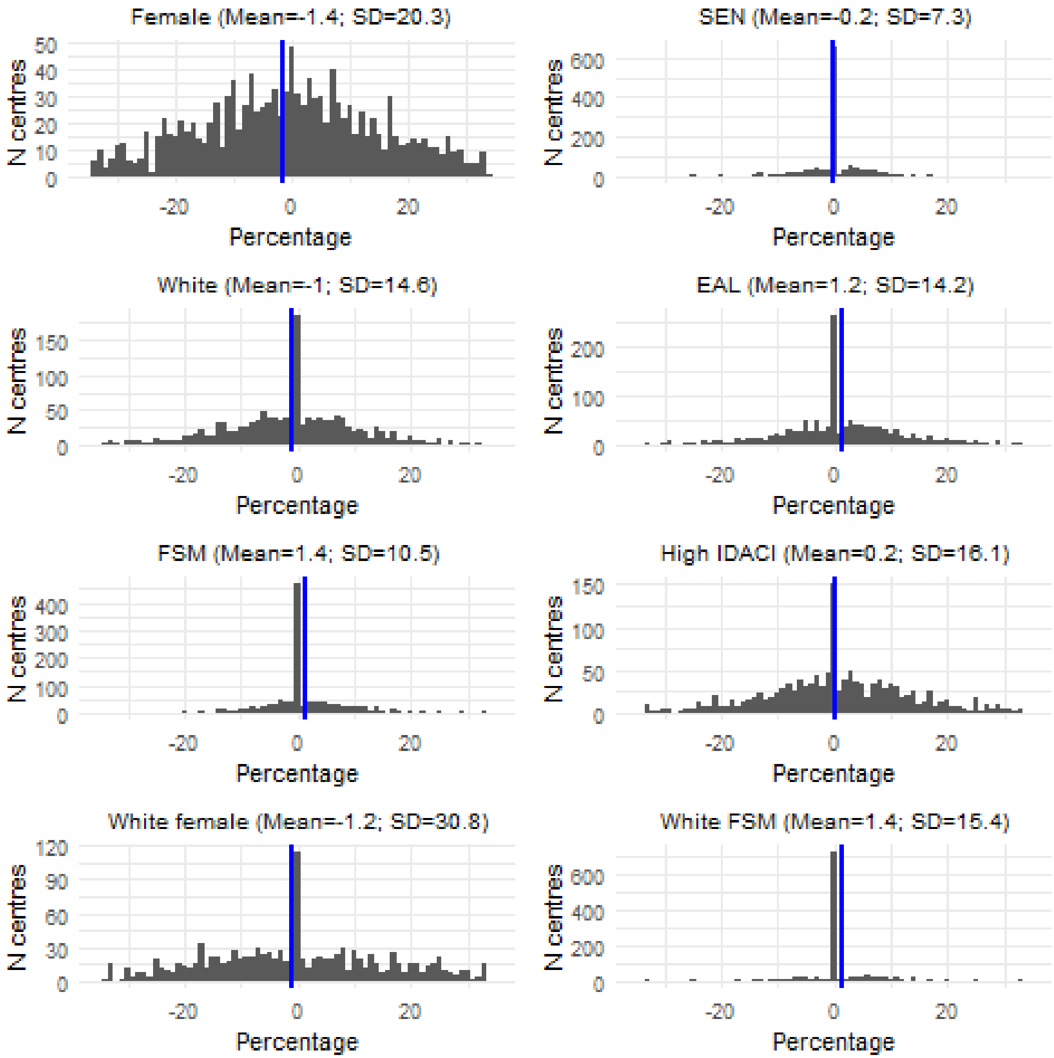
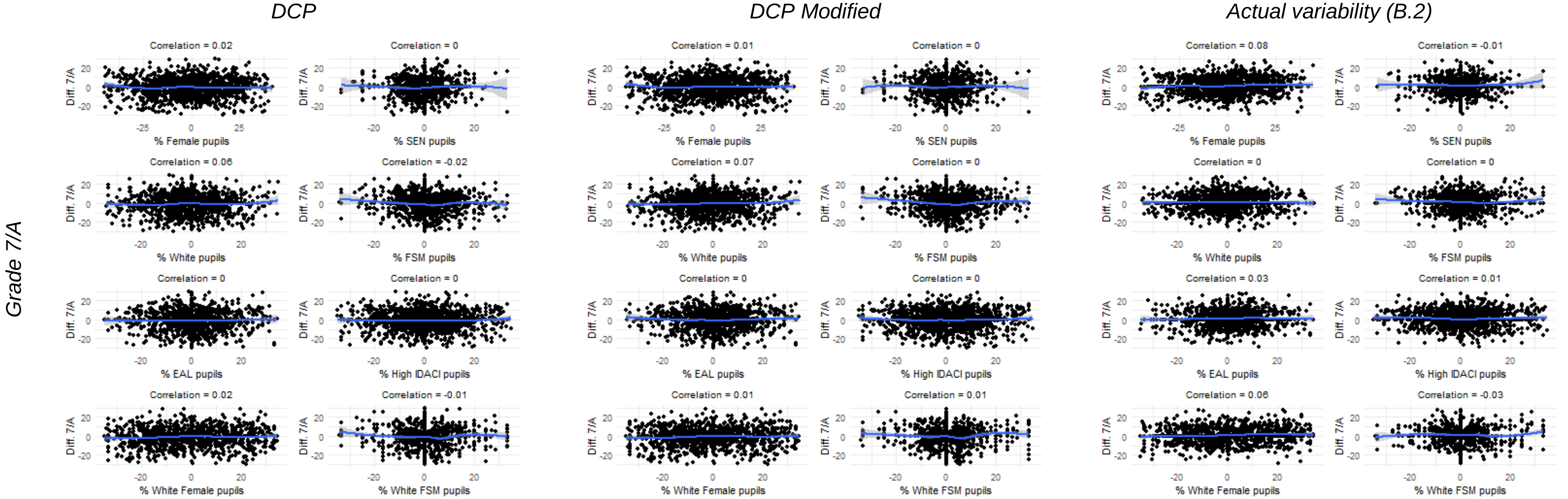


Figure G.18. Correlation between demographic/socio-economic factors and accuracy of predictions – Biology



Grade 4/C

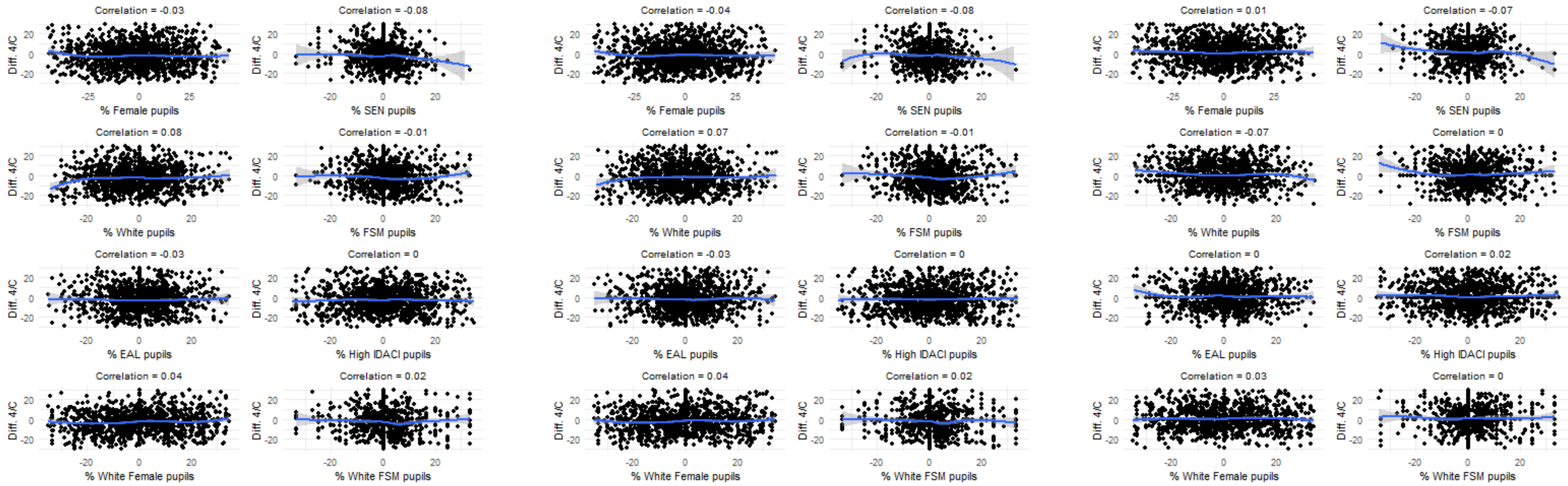

Correlation = -0.03
Diff. 4/C
% Female pupils
Correlation = -0.08
% SEN pupils
Correlation = 0.08
% White pupils
Correlation = -0.01
% FSM pupils
Correlation = -0.03
% EAL pupils
Correlation = 0
% High IDACI pupils
Correlation = 0.04
% White Female pupils
Correlation = 0.02
% White FSM pupils
Correlation = -0.04
% Female pupils
Correlation = -0.08
% SEN pupils
Correlation = 0.07
% White pupils
Correlation = -0.01
% FSM pupils
Correlation = -0.03
% EAL pupils
Correlation = 0
% High IDACI pupils
Correlation = 0.04
% White Female pupils
Correlation = 0.02
% White FSM pupils
Correlation = 0.01
% Female pupils
Correlation = -0.07
% SEN pupils
Correlation = -0.07
% White pupils
Correlation = 0
% FSM pupils
Correlation = 0
% EAL pupils
Correlation = 0.02
% High IDACI pupils
Correlation = 0.03
% White Female pupils
Correlation = 0
% White FSM pupils

Table G.6. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Biology

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	-20.22 ***	-39.41 ***	-22.82 ***	-36.00 ***	4.18	49.98 ***
	(3.93)	(5.49)	(4.00)	(5.82)	(3.21)	(4.80)
Perc.Female	0.01	-0.07 **	0.00	-0.08 **	0.02	-0.01
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
Perc.SEN	0.00	-0.19 ***	-0.01	-0.20 ***	-0.01	-0.12 *
	(0.04)	(0.06)	(0.04)	(0.06)	(0.03)	(0.05)
Perc.WHITE	0.05 *	0.06 *	0.05 *	0.06	0.00	-0.08 **
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Perc.FSM	-0.02	-0.03	-0.01	-0.03	0.02	-0.01
	(0.03)	(0.05)	(0.03)	(0.05)	(0.03)	(0.04)
Perc.EAL	0.02	-0.01	0.02	-0.02	0.02	-0.03
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
IDACI.high	0.02	0.01	0.01	-0.00	0.01	0.02
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
IDACI.med	0.01	0.01	0.00	-0.00	0.01	0.02
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
WHITE_Female	0.00	0.04 *	-0.00	0.05 **	0.01	0.03
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
WHITE_FSM	0.00	0.03	0.01	0.03	-0.03	-0.00
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
mean.GCSE	0.30 ***	0.57 ***	0.36 ***	0.55 ***	-0.04	-0.74 ***
	(0.06)	(0.08)	(0.06)	(0.09)	(0.05)	(0.07)
Ncands_2019	-0.06 ***	-0.07 **	-0.08 ***	-0.11 ***	-0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
N	1389	1389	1389	1389	1389	1389
R2	0.03	0.05	0.03	0.05	0.01	0.08

***** p < 0.001; ** p < 0.01; * p < 0.05.**

A level French

Figure G.19. Year-on-year variability in outcomes – French

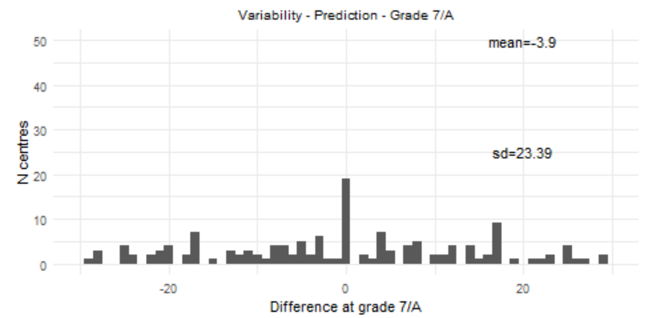
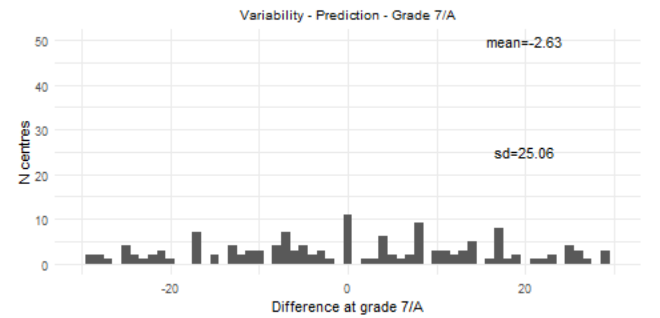
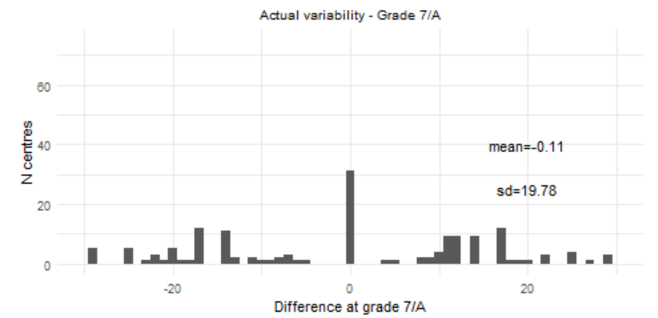
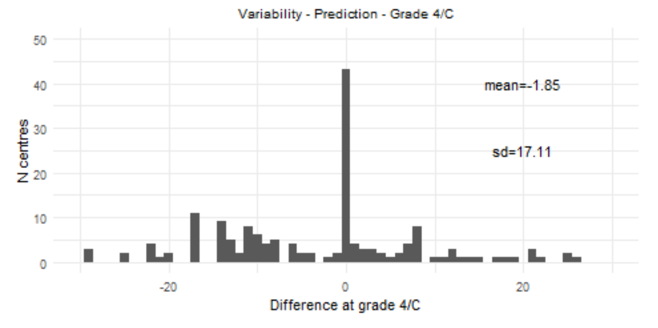
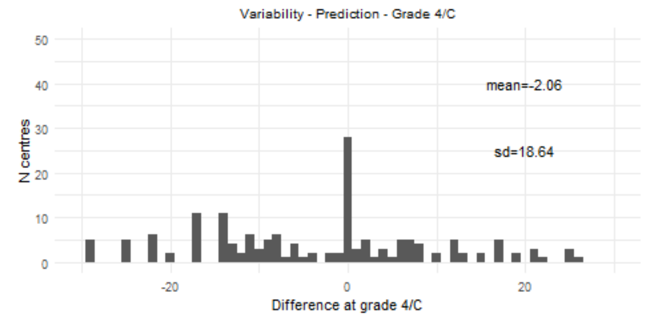
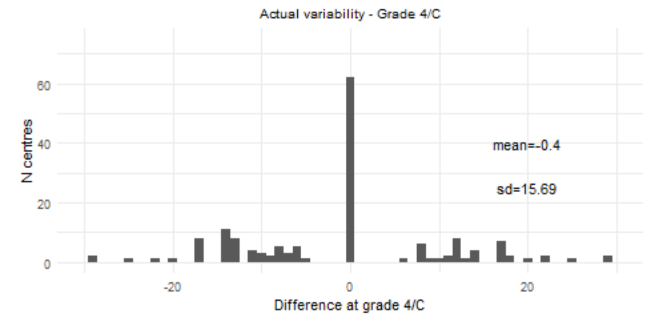
<table>
<tr><th></th><th>DCP</th><th>DCP Modified</th><th>Actual variability (B.2)</th></tr>
<tr><td>Grade 7/A</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
<tr><td>Grade 4/C</td>
<td>
</td>
<td>
</td>
<td>
</td></tr>
</table>

Figure G.20. Year-on-year variability in demographic and socio-economic centre composition – French

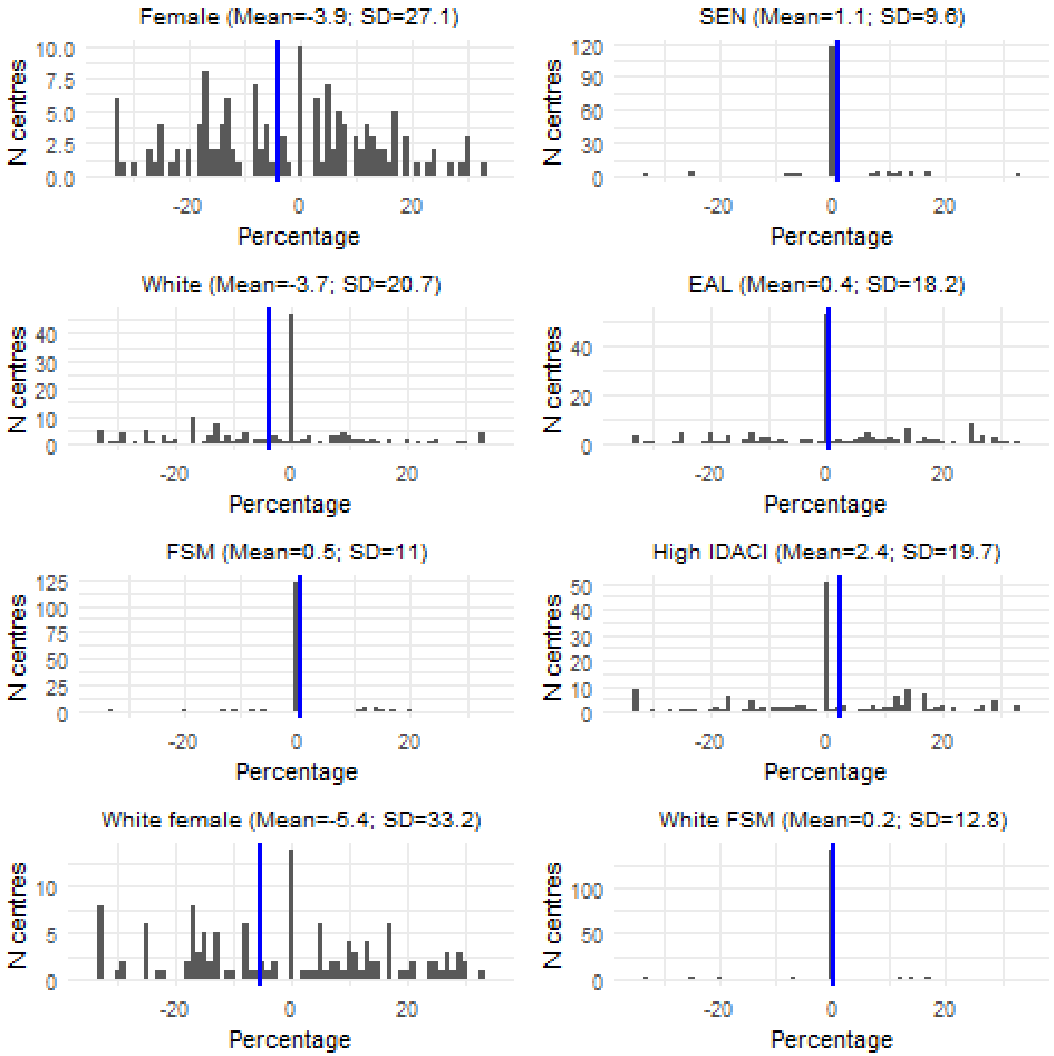
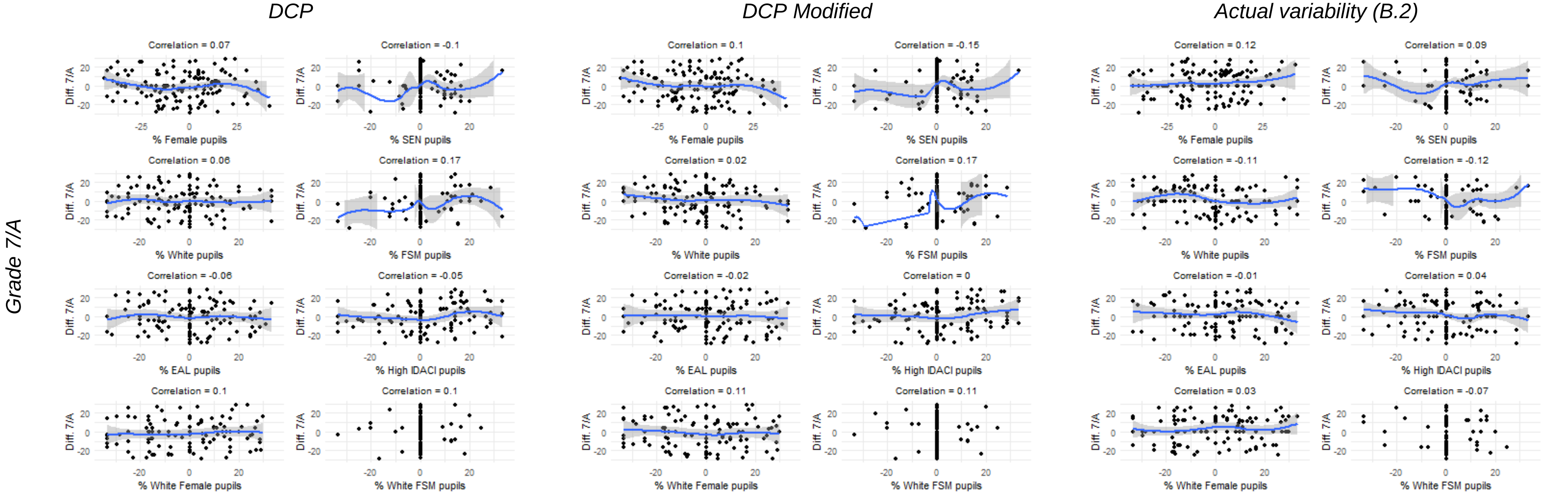


Figure G.21. Correlation between demographic/socio-economic factors and accuracy of predictions – French



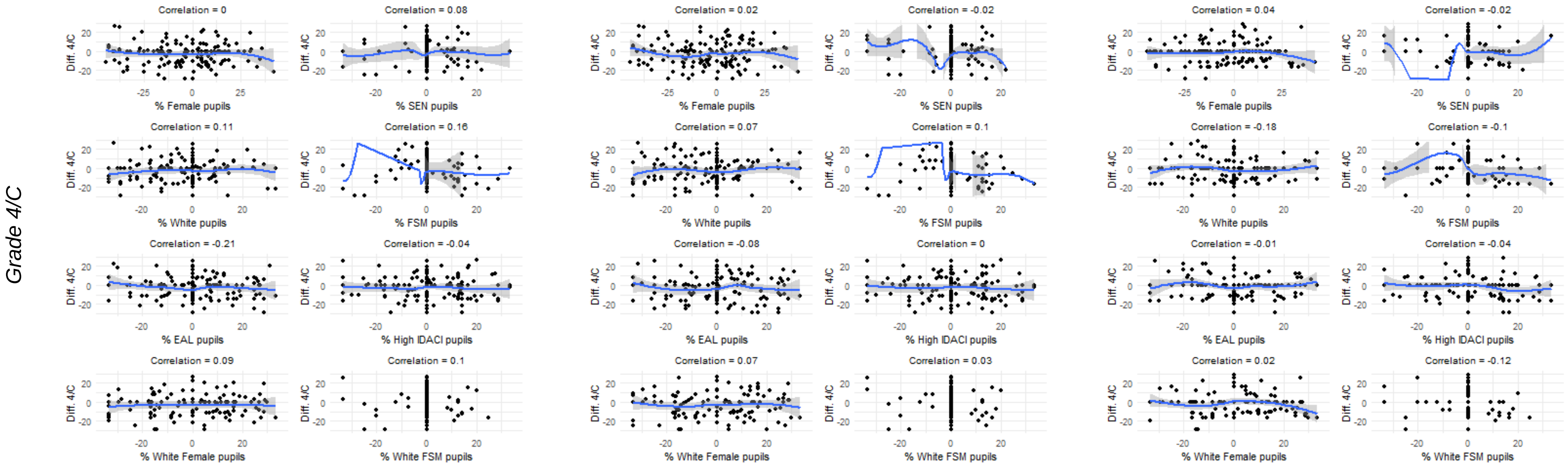

Grade 4/C
Correlation = 0
Correlation = 0.08
Correlation = 0.02
Correlation = -0.02
Correlation = 0.04
Correlation = -0.02
Correlation = 0.11
Correlation = 0.16
Correlation = 0.07
Correlation = 0.1
Correlation = -0.18
Correlation = -0.1
Correlation = -0.21
Correlation = -0.04
Correlation = -0.08
Correlation = 0
Correlation = -0.01
Correlation = -0.04
Correlation = 0.09
Correlation = 0.1
Correlation = 0.07
Correlation = 0.03
Correlation = 0.02
Correlation = -0.12
Diff. 4/C
% Female pupils
% SEN pupils
% White pupils
% FSM pupils
% EAL pupils
% High IDACI pupils
% White Female pupils
% White FSM pupils

Table G.7. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – French

	DCP		**DCP Modified**		**Actual (B.2)**	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	-75.47 ***	-23.26	-70.22 ***	-34.57 *	-2.79	4.05
	(17.60)	(13.29)	(18.95)	(14.86)	(15.64)	(12.33)
Perc.Female	0.02	-0.08	0.03	-0.04	0.15	-0.10
	(0.14)	(0.10)	(0.15)	(0.12)	(0.12)	(0.10)
Perc.SEN	-0.19	0.14	-0.31	-0.02	0.34 *	0.13
	(0.18)	(0.14)	(0.20)	(0.15)	(0.16)	(0.13)
Perc.WHITE	0.01	0.02	-0.03	0.04	-0.06	-0.17 **
	(0.09)	(0.07)	(0.10)	(0.08)	(0.08)	(0.07)
Perc.FSM	0.27	0.20	0.19	0.16	-0.16	-0.07
	(0.22)	(0.17)	(0.24)	(0.19)	(0.20)	(0.16)
Perc.EAL	-0.03	-0.17 *	-0.00	-0.04	-0.04	-0.13
	(0.10)	(0.08)	(0.11)	(0.09)	(0.09)	(0.07)
IDACI.high	-0.08	-0.01	-0.05	-0.01	-0.06	-0.10
	(0.10)	(0.08)	(0.11)	(0.08)	(0.09)	(0.07)
IDACI.med	-0.01	-0.03	-0.03	-0.05	-0.07	-0.04
	(0.07)	(0.05)	(0.07)	(0.06)	(0.06)	(0.05)
WHITE_Female	0.05	0.09	0.06	0.05	-0.10	0.10
	(0.11)	(0.09)	(0.12)	(0.10)	(0.10)	(0.08)
WHITE_FSM	0.09	0.03	0.20	-0.04	0.03	0.03
	(0.19)	(0.15)	(0.21)	(0.16)	(0.17)	(0.14)
mean.GCSE	1.11 ***	0.30	1.09 ***	0.47 *	0.13	-0.09
	(0.26)	(0.19)	(0.28)	(0.22)	(0.23)	(0.18)
Ncands_2019	-0.69	0.03	-0.99	-0.08	-0.73	0.21
	(0.54)	(0.41)	(0.59)	(0.46)	(0.48)	(0.38)
N	171	171	171	171	171	171
R2	0.16	0.10	0.15	0.06	0.07	0.08

***** p < 0.001; ** p < 0.01; * p < 0.05.**

A level Religious studies

Figure G.22. Year-on-year variability in outcomes – Religious studies

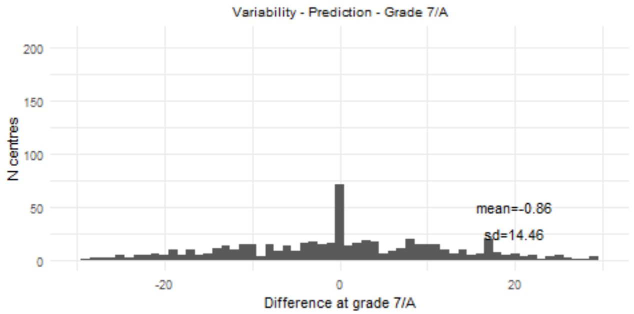
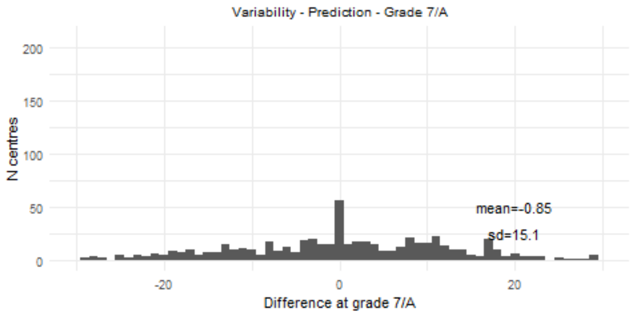
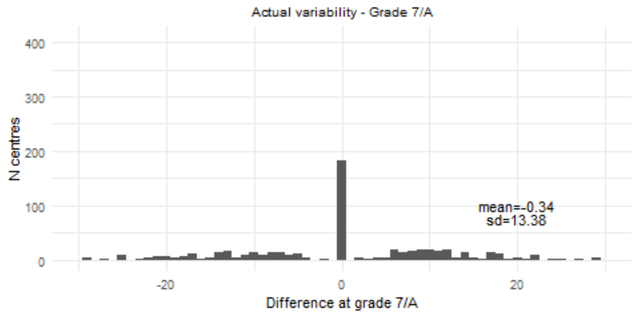
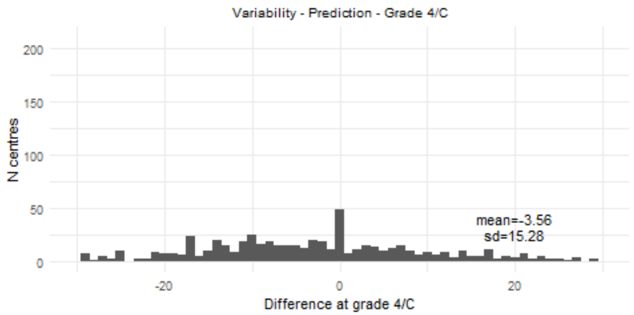
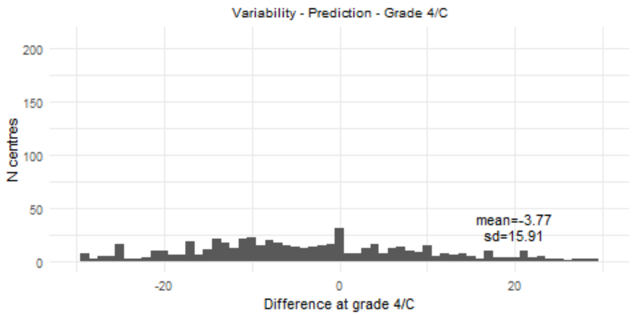
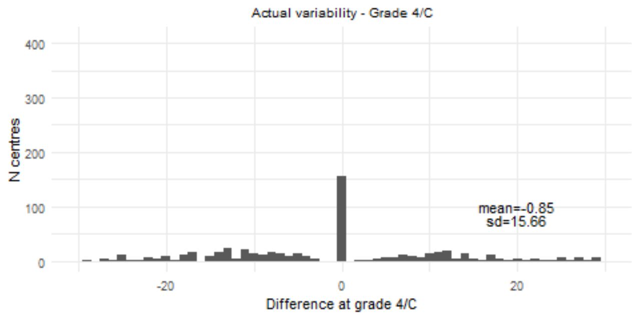
	DCP	*DCP Modified*	*Actual variability (B.2)*
Grade 7/A			
Grade 4/C			

Figure G.23. Year-on-year variability in demographic and socio-economic centre composition – Religious studies

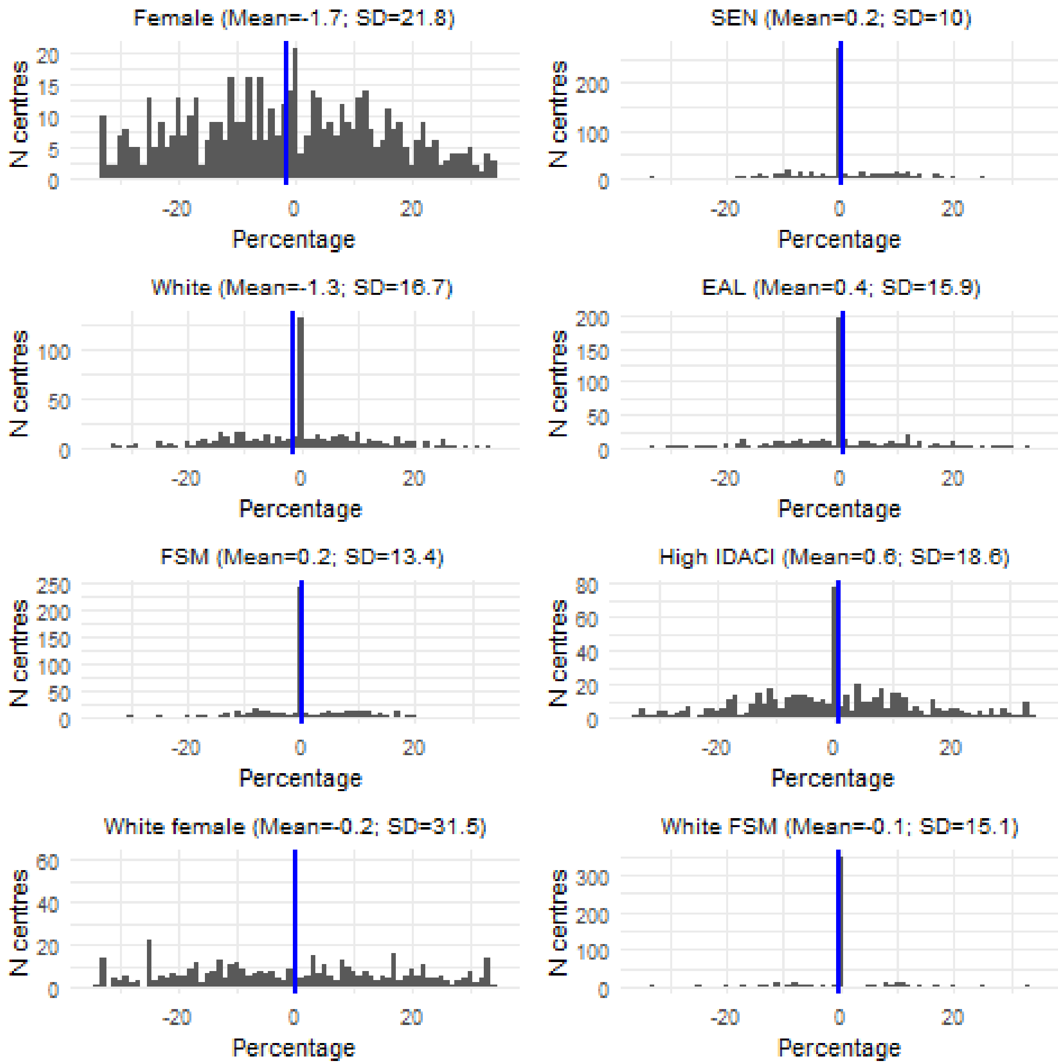
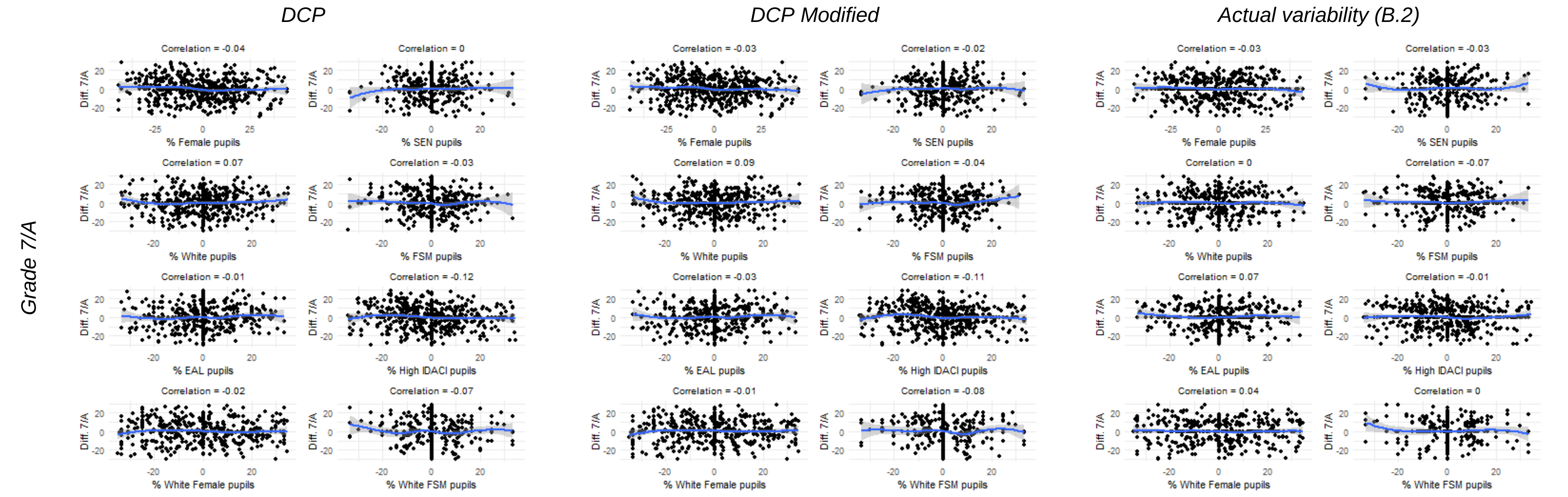


Figure G.24. Correlation between demographic/socio-economic factors and accuracy of predictions – Religious studies



Grade 4/C

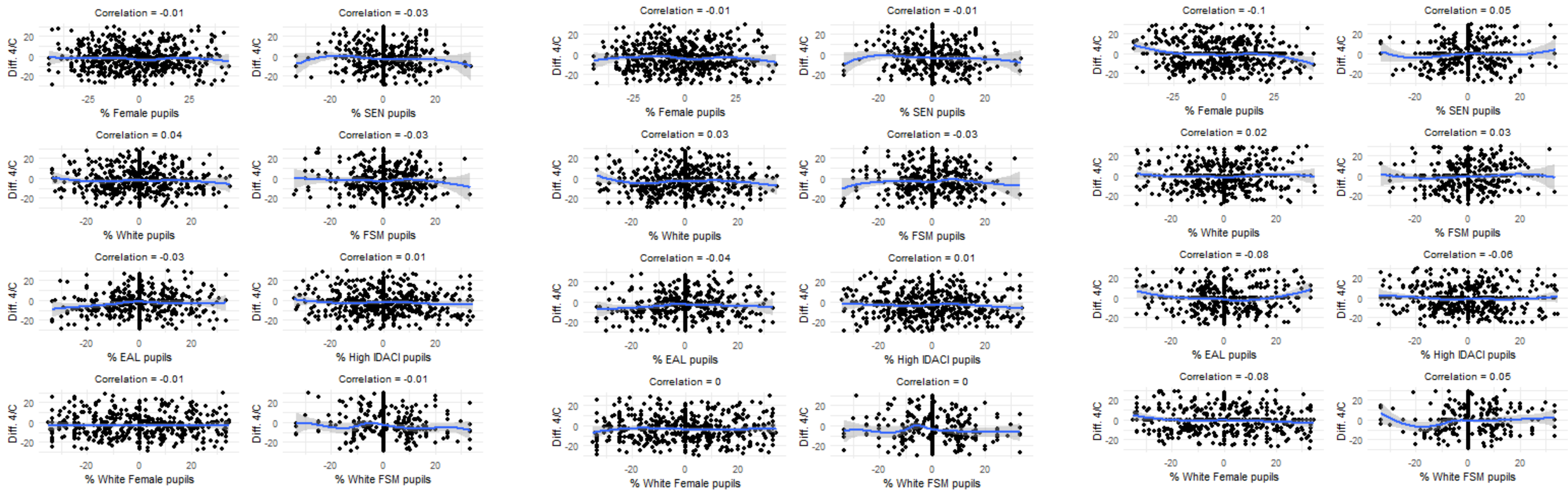

Correlation = -0.01
% Female pupils
Correlation = -0.03
% SEN pupils
Correlation = 0.04
% White pupils
Correlation = -0.03
% FSM pupils
Correlation = -0.03
% EAL pupils
Correlation = 0.01
% High IDACI pupils
Correlation = -0.01
% White Female pupils
Correlation = -0.01
% White FSM pupils
Correlation = -0.01
% Female pupils
Correlation = -0.01
% SEN pupils
Correlation = 0.03
% White pupils
Correlation = -0.03
% FSM pupils
Correlation = -0.04
% EAL pupils
Correlation = 0.01
% High IDACI pupils
Correlation = 0
% White Female pupils
Correlation = 0
% White FSM pupils
Correlation = -0.1
% Female pupils
Correlation = 0.05
% SEN pupils
Correlation = 0.02
% White pupils
Correlation = 0.03
% FSM pupils
Correlation = -0.08
% EAL pupils
Correlation = -0.06
% High IDACI pupils
Correlation = -0.08
% White Female pupils
Correlation = 0.05
% White FSM pupils
Diff. 4/C

Table G.8. Regression estimates for the impact of demographic and socio-economic centre composition on accuracy – Religious studies

	DCP		DCP Modified		Actual (B.2)	
	Var. 7/A+	**Var. 4/C+**	**Var. 7/A+**	**Var. 4/C+**	**Act. 7/A+**	**Act. 4/C+**
(Intercept)	-32.92 ***	-32.91 ***	-30.81 ***	-28.44 ***	7.43	16.13 *
	(5.89)	(6.34)	(6.18)	(6.65)	(5.14)	(6.27)
Perc.Female	-0.06	-0.02	-0.05	-0.02	-0.05	-0.05
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Perc.SEN	0.01	-0.03	-0.01	0.00	-0.08	0.06
	(0.06)	(0.06)	(0.06)	(0.07)	(0.05)	(0.06)
Perc.WHITE	0.05	0.04	0.07	0.02	0.01	-0.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
Perc.FSM	0.05	-0.05	0.05	-0.05	-0.10 *	-0.03
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)
Perc.EAL	0.03	-0.00	0.02	-0.02	0.04	-0.10 *
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
IDACI.high	-0.09 *	0.01	-0.07	0.01	-0.04	-0.08 *
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)
IDACI.med	-0.04	-0.04	-0.02	-0.03	-0.05	-0.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
WHITE_Female	0.02	-0.00	0.02	0.00	0.04	-0.01
	(0.02)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
WHITE_FSM	-0.09	0.01	-0.11 *	0.02	0.02	0.07
	(0.05)	(0.05)	(0.05)	(0.06)	(0.04)	(0.05)
mean.GCSE	0.57 ***	0.47 ***	0.53 ***	0.40 ***	-0.15	-0.32 **
	(0.10)	(0.11)	(0.10)	(0.11)	(0.09)	(0.11)
Ncands_2019	-0.09	0.12	-0.11	0.10	0.08	0.10
	(0.08)	(0.09)	(0.09)	(0.10)	(0.07)	(0.09)
N	586	586	586	586	586	586
R2	0.08	0.04	0.07	0.03	0.03	0.05

***** p < 0.001; ** p < 0.01; * p < 0.05.**

Annex H: Harmonic mean

The harmonic mean is used in the standardisation process to calculate the size of the centre's entry in a subject across the current year and the historical data. There are important differences between the harmonic mean used here and the arithmetic mean, which is the more typical calculation implied when reference is made to the 'mean' or 'average'.

The general equation for calculating the harmonic mean, often denoted as H, is shown below:

$$H = \frac{n}{\left(\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}\right)}$$

where X represents the values of which the harmonic mean is being calculated and n is the number of those values. For the case used here, the harmonic mean is calculated based on 2 data points – the number of students in the current year sitting the subject with the centre and the number of students in the historical data sitting the subject with the same centre. This means the calculation of the harmonic mean of students can be written as:

$$\bar{n}_H = \frac{2}{\left(\frac{1}{n_{\text{cur}}} + \frac{1}{n_{\text{hist}}}\right)}$$

$$= \left(\frac{0.5}{n_{\text{cur}}} + \frac{0.5}{n_{\text{hist}}}\right)^{-1}$$

Use of the harmonic mean to characterise the size of a centre's entry ensures that an inappropriate weight is not put on the statistical prediction where one of the numbers – either n_{cur} or n_{hist} – is small. With the arithmetic mean, the two values being used in the calculation have equal weight when they are summed together. However, when calculating the harmonic mean the use of the reciprocal of the number ensures that there is greater weight put on the smaller of the two values.

This helps ensure that, for example, a centre with very few students this year, but a large number in the historical data is not classed as a 'large' centre and inappropriately adjusted using the statistical model. This may be the case were the arithmetic mean used to characterise the size of the entry. It would be just as inappropriate to base a prediction for a large number of students this year on the performance of a small cohort in the historical data as it would be to use a prediction for a small number of students this year.

This effect is shown in Figure H.1 for the example of where there are 3 students in the current cohort. This plot shows that, as the number of students making up the historical data changes, the arithmetic mean changes linearly. However, the relationship with the harmonic mean is non-linear, with the value being weighted towards the lower value.

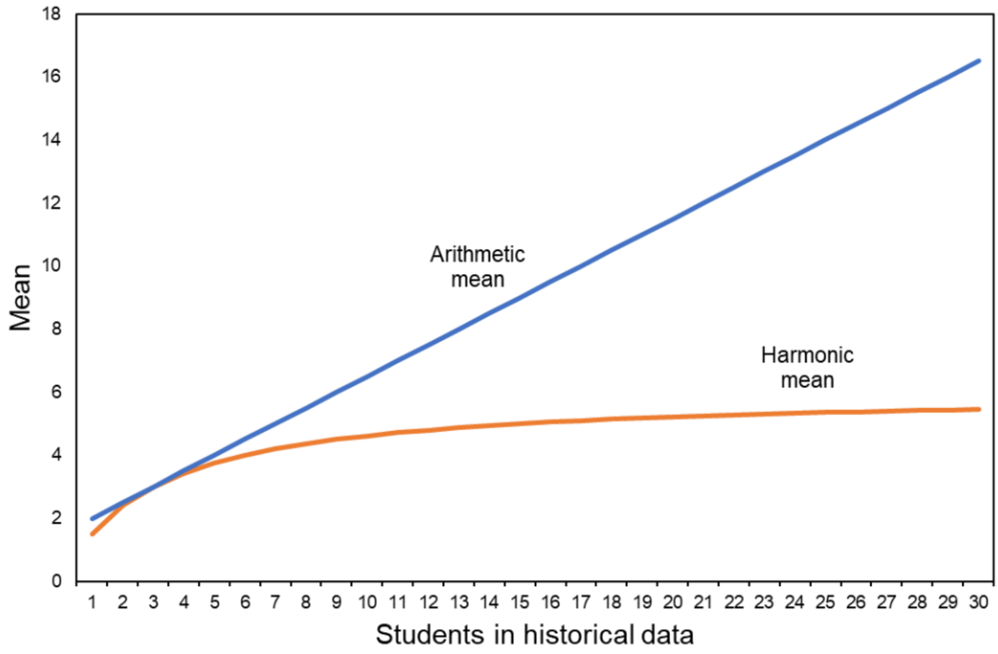


Figure H.1 Variation of arithmetic mean and harmonic mean for different values of n_{hist} with n_{cur} = 3

Figure H.1 shows that, the value of arithmetic mean increases unbounded with the number of students in the historical data. This is not the case with the harmonic mean. Indeed, for the example above, it can be shown that the value of harmonic mean is limited to 6; twice the value of n_{cur}. This can be shown to be a fundamental limit as demonstrated below:

$$\bar{n}_H = \left(\frac{0.5}{n_{\text{cur}}} + \frac{0.5}{\infty}\right)^{-1}$$

$$= \left(\frac{0.5}{n_{\text{cur}}} + 0\right)^{-1}$$

$$= 2n_{\text{cur}}$$

Annex I: Centres with a small entry in a subject in 2019

Table I.1. Centre-level analysis for Phase 1 reform GCSE subjects

		Based on data from year - 1					Based on data from across year - 1 and year - 2				
Subject	Number of centres in 2019	Percentage of centres without history	Percentage of centres with…				Percentage of centres without history	Percentage of centres with…			
			$\bar{n}_H \leq 5$	$\bar{n}_H \leq 10$	$\bar{n}_H \leq 15$	$\bar{n}_H \leq 20$		$\bar{n}_H \leq 5$	$\bar{n}_H \leq 10$	$\bar{n}_H \leq 15$	$\bar{n}_H \leq 20$
English Language	5351	4.7	5.9	11.7	16.3	19.2	3.8	5.0	10.2	14.1	17.4
English Literature	4240	6.7	5.4	9.2	11.7	13.5	5.5	5.1	8.3	11.0	12.6
Mathematics	5428	4.2	6.4	13.1	17.4	20.7	3.5	5.1	11.0	15.1	18.6

Table I.2. Candidate-level analysis for Phase 1 reform GCSE subjects

		Based on data from year - 1					Based on data from across year - 1 and year - 2				
Subject	Number of candidates in 2019	Percentage of candidates from centres without history	Percentage of candidates from centres with…				Percentage of candidates from centres without history	Percentage of candidates from centres with…			
			$\bar{n}_H \leq 5$	$\bar{n}_H \leq 10$	$\bar{n}_H \leq 15$	$\bar{n}_H \leq 20$		$\bar{n}_H \leq 5$	$\bar{n}_H \leq 10$	$\bar{n}_H \leq 15$	$\bar{n}_H \leq 20$
English Language	706,620	0.9	0.2	0.7	1.1	1.6	0.8	0.2	0.5	0.9	1.2
English Literature	546,638	1.3	0.3	0.9	1.2	1.5	1.2	0.2	0.3	0.5	0.7
Mathematics	719,874	0.7	0.2	0.7	1.1	1.6	0.7	0.2	0.5	0.8	1.2

Table I.3. Centre and candidates-level analysis for Phase 2 reform GCSE subjects

	Centre-level analysis						Candidate-level analysis					
Subject	Number of centres in 2019	Percentage of centres without history	Percentage of centres with $\widehat{n_H} \leq$ x based on data from year – 1				Number of candidates in 2019	Percentage of candidates from centres without history	Percentage of candidates from centres with $\widehat{n_H} \leq$ x based on data from year – 1			
			$x = 5$	$x = 10$	$x = 15$	$x = 20$			$x = 5$	$x = 10$	$x = 15$	$x = 20$
Art & Design	4,366	4.4	7.6	13.9	19.6	26.4	182,216	1.3	0.7	1.9	3.9	7.0
Biology	3,887	9.3	4.5	9.3	14.8	21.3	165,072	2.7	0.6	1.6	3.5	6.4
Chemistry	3,528	8.5	3.3	6.5	10.8	17.5	159,038	2.8	0.3	1.0	2.4	5.3
Citizenship Studies	527	32.3	4.9	13.5	21.3	30.6	19,644	20.1	1.8	3.9	6.9	11.7
Classical Greek	201	24.9	41.3	64.2	69.2	72.6	1,143	8.1	25.1	56.8	66.8	77.0
Computing	2,972	13.3	3.7	9.8	21.9	36.5	77,419	8.5	0.7	3.2	10.0	21.1
Dance	735	23.8	8.0	30.1	55.4	66.3	9,274	17.6	2.3	18.1	45.4	61.3
Drama	2,550	8.6	2.7	12.4	29.8	47.8	57,560	5.9	0.6	4.4	15.4	30.1
Food Prep and Nutrition	2,001	9.5	3.7	12.6	30.0	51.0	44,921	6.7	0.8	4.4	15.2	32.7
French	3,340	7.3	6.5	14.5	23.4	32.5	121,756	3.8	1.3	3.6	7.2	11.9
Geography	3,832	4.1	2.5	5.1	7.6	10.6	250,986	1.0	0.1	0.5	1.0	1.9
German	1,676	21.1	11.3	19.7	28.9	38.1	40,655	6.2	2.1	5.1	10.7	17.9
History	3,848	5.6	4.7	7.5	10.0	12.0	260,468	1.1	0.2	0.6	1.1	1.7

Latin	642	21.3	14.3	33.5	46.3	57.9	9,355	6.7	3.5	13.5	25.1	39.7
Music	2623	8.4	12.2	36.9	64.0	78.5	34,049	5.2	3.8	20.1	47.5	67.7
Physical Education	2732	8.2	4.9	11.2	19.7	30.3	79,942	4.9	0.9	2.7	6.7	13.2
Physics	3508	8.5	3.3	6.2	10.6	17.1	157,787	2.7	0.4	1.0	2.3	5.2
Religious Studies	3191	10.6	5.0	10.2	17.6	24.3	249,325	3.3	0.5	1.1	2.4	4.2
Spanish	2912	13.0	10.8	18.2	25.9	33.8	95,860	5.8	2.3	5.1	9.0	14.0

Table I.4. Centre-level analysis for A level

		Based on data from year – 1					Based on data from across year – 1 to year - 3				
Subject	Number of centres in 2019	Percentage of centres without history	Percentage of centres with $\widehat{n_H} \leq$ x				Percentage of centres without history	Percentage of centres with $\widehat{n_H} \leq$ x			
			$x = 5$	$x = 10$	$x = 15$	$x = 20$		$x = 5$	$x = 10$	$x = 15$	$x = 20$
Accounting	195	23.6	29.7	47.7	56.4	61.5	16.9	27.2	43.6	53.8	59.5
Ancient History	77	46.8	22.1	37.7	40.3	45.5	40.3	18.2	36.4	42.9	45.5
Arabic	305	56.4	36.4	40.7	42.3	43.3	35.1	54.4	58.7	62.0	63.3
Art & Design	2,335	5.4	23.3	50.3	66.6	75.9	2.5	14.0	34.9	52.6	64.2
Bengali	24	79.2	20.8	20.8	20.8	20.8	66.7	33.3	33.3	33.3	33.3
Biblical Hebrew	13	46.2	30.8	46.2	53.8	53.8	15.4	46.2	69.2	76.9	84.6
Biology	2,577	4.1	12.3	30.7	45.0	57.2	2.5	6.6	19.0	31.1	42.3
Business Studies	1,775	7.8	15.3	39.2	57.6	69.1	4.5	10.8	26.1	43.2	55.3

Chemistry	2,543	4.7	16.8	38.3	53.6	64.3	2.9	9.7	25.5	39.4	49.7
Chinese	395	28.6	37.2	53.9	64.8	68.6	19.0	37.2	52.9	66.6	72.4
Classical Civilisation	403	26.1	39.2	55.6	66.7	69.5	14.9	31.0	55.1	67.5	74.7
Classical Greek	91	29.7	65.9	69.2	69.2	70.3	16.5	70.3	82.4	82.4	82.4
Computing	1,300	20.8	33.9	62.3	71.2	74.4	16.5	25.5	54.2	69.0	74.8
D&T: Product Design	1,260	12.0	39.2	72.4	82.4	85.6	5.4	24.8	58.2	77.5	86.2
Dance	210	18.6	51.4	74.8	79.5	81.0	7.1	40.0	73.3	85.7	89.0
Drama & Theatre Studies	1,241	14.3	35.8	69.5	78.8	81.9	3.5	25.5	62.6	79.4	86.8
Dutch	26	38.5	61.5	61.5	61.5	61.5	26.9	73.1	73.1	73.1	73.1
Economics	1,619	6.5	13.7	34.8	53.6	65.2	4.0	9.4	22.7	37.4	51.4
Electronics	55	10.9	34.5	63.6	72.7	78.2	5.5	16.4	41.8	63.6	67.3
English Language	950	9.3	15.3	41.9	62.1	72.2	3.6	9.7	25.4	46.2	60.5
English Language & Literature	577	15.9	15.6	44.5	61.4	70.0	10.4	11.1	29.6	48.9	62.4
English Literature	2,344	6.1	15.1	39.3	58.3	70.8	2.5	8.2	24.8	41.7	56.4
Environmental Studies	65	40.0	9.2	29.2	38.5	47.7	35.4	13.8	20.0	35.4	38.5
Film Studies	513	22.8	17.3	45.8	60.0	66.5	16.8	9.6	29.8	50.5	63.7

French	1,500	17.7	54.9	74.1	78.6	80.4	5.5	47.3	74.5	85.9	89.6
Further mathematics	1,857	12.0	44.3	70.9	78.9	82.1	5.3	33.5	63.1	76.8	83.6
Geography	2,134	6.8	21.3	47.1	62.3	72.8	2.7	13.4	33.8	48.7	60.2
Geology	98	14.3	16.3	44.9	57.1	66.3	6.1	9.2	27.6	48.0	61.2
German	818	29.6	57.5	68.2	69.6	69.9	10.3	58.7	80.9	86.9	88.6
Greek (Modern)	119	74.8	23.5	25.2	25.2	25.2	55.5	41.2	43.7	43.7	44.5
Gujarati	10	70.0	30.0	30.0	30.0	30.0	70.0	10.0	30.0	30.0	30.0
History	2,446	5.1	15.4	37.2	53.4	64.8	2.3	8.7	23.6	38.1	49.7
Italian	340	47.1	45.6	51.8	52.6	52.9	30.6	57.1	65.3	68.5	69.1
Japanese	118	61.0	36.4	38.1	39.0	39.0	39.0	57.6	60.2	60.2	61.0
Latin	319	21.6	61.8	75.2	77.7	77.7	8.8	58.6	82.1	87.1	90.6
Law	597	14.6	18.3	40.2	53.9	61.0	9.5	15.6	28.8	43.4	55.1
Mathematics	2,730	5.3	11.8	27.4	39.0	48.8	3.4	7.5	18.7	29.0	37.3
Media Studies	1,024	7.4	14.3	42.9	65.3	76.3	4.5	7.4	26.3	46.4	62.1
Modern Hebrew	17	47.1	35.3	52.9	52.9	52.9	29.4	41.2	58.8	70.6	70.6
Music	1,080	17.4	61.3	75.5	79.3	80.9	6.2	54.3	79.6	86.7	89.1
Panjabi	57	54.4	35.1	42.1	43.9	43.9	36.8	47.4	59.6	59.6	61.4

Persian	98	68.4	28.6	29.6	30.6	30.6	53.1	39.8	44.9	44.9	44.9
Philosophy	291	24.7	24.4	51.2	61.9	67.4	22.0	22.3	42.6	55.0	64.9
Physics	2,432	6.1	22.8	49.7	64.8	74.8	2.9	15.0	35.1	51.8	63.0
Polish	413	54.7	41.4	43.8	44.6	45.0	32.0	56.7	65.1	66.6	67.3
Politics	1,340	14.7	16.4	41.5	62.4	72.2	8.5	12.9	29.4	48.1	61.0
Portuguese	230	60.4	33.5	38.3	39.1	39.6	40.4	50.0	57.0	58.7	59.6
Psychology	2,367	5.1	9.6	25.7	40.5	54.1	3.3	6.0	16.1	27.1	37.3
Religious Studies	1,563	8.8	24.9	58.0	73.8	82.7	2.9	14.6	38.8	60.0	73.8
Russian	286	42.7	47.9	55.9	56.3	57.0	25.9	57.7	67.8	72.4	72.7
Sociology	1,715	9.1	12.7	31.4	49.8	64.3	5.4	7.6	19.7	33.6	46.6
Spanish	1,414	22.0	46.9	69.6	74.1	75.6	6.9	42.6	70.6	83.5	88.0
Sport & PE	1,265	11.3	38.3	71.7	82.5	84.7	4.0	24.8	59.1	78.8	87.7
Statistics	52	48.1	9.6	23.1	32.7	36.5	42.3	9.6	19.2	25.0	38.5
Turkish	213	57.3	36.2	39.4	41.3	42.3	39.0	47.4	53.5	57.7	58.7
Urdu	83	53.0	22.9	39.8	42.2	42.2	33.7	37.3	49.4	59.0	61.4

Table I.5. Candidate-level analysis for A level

		Based on data from year – 1					Based on data from across year – 1 to year - 3				
Subject	Number of candidates in 2019	Percentage of candidates from centres without history	Percentage of candidates from centres with $\widehat{n_H} \leq$ x				Percentage of candidates from centres without history	Percentage of candidates from centres with $\widehat{n_H} \leq$ x			
			$x = 5$	$x = 10$	$x = 15$	$x = 20$		$x = 5$	$x = 10$	$x = 15$	$x = 20$
Accounting	2,215	6.7	8.7	21.3	31.8	40.0	5.3	5.0	13.9	21.7	28.1
Ancient History	681	41.0	7.2	20.3	24.2	33.2	39.2	3.7	15.0	20.6	23.5
Arabic	765	32.7	30.3	45.1	53.9	59.7	19.7	35.8	45.5	58.0	65.0
Art & Design	38,597	1.7	5.1	18.1	30.8	40.9	1.0	1.8	8.4	17.6	26.0
Bengali	28	75.0	25.0	25.0	25.0	25.0	57.1	42.9	42.9	42.9	42.9
Biblical Hebrew	66	39.4	21.2	42.4	60.6	60.6	12.1	31.8	59.1	69.7	87.9
Biology	63,580	1.3	2.2	8.5	16.4	25.8	1.0	0.7	3.4	7.9	13.7
Business Studies	30,236	2.7	3.7	15.6	30.1	42.7	1.9	1.6	6.8	16.2	25.2
Chemistry	54,868	1.6	3.1	11.5	21.2	30.6	1.0	1.1	5.3	11.1	17.1
Chinese	2,086	7.9	17.4	42.7	65.0	76.0	4.5	12.0	26.7	48.2	61.2
Classical Civilisation	2,893	20.2	18.0	34.0	53.1	59.6	11.9	9.2	26.1	40.9	52.9
Classical Greek	210	20.0	62.9	71.9	71.9	80.0	10.0	55.7	81.9	81.9	81.9
Computing	10,435	12.2	16.5	43.6	57.7	64.4	9.8	9.6	31.7	48.7	58.1
D&T: Product Design	9,214	7.1	20.3	57.6	74.7	82.7	3.2	8.4	33.2	57.2	72.3

Dance	1,016	12.1	34.0	68.6	80.9	86.0	3.9	20.4	54.4	75.0	83.3
Drama & Theatre Studies	8,970	7.9	17.8	51.6	67.6	74.7	2.1	9.1	35.7	55.0	67.3
Dutch	33	33.3	66.7	66.7	66.7	66.7	24.2	75.8	75.8	75.8	75.8
Economics	29,740	2.0	2.8	12.3	25.8	37.2	1.4	1.1	5.2	12.4	21.7
Electronics	588	4.6	12.6	34.0	46.9	56.3	0.5	3.2	14.8	31.5	35.4
English Language	13,560	4.7	4.8	19.4	36.8	48.3	1.7	1.2	7.2	19.1	30.7
English Language & Literature	7,425	7.5	4.3	22.1	38.3	49.4	5.0	2.4	10.0	22.4	34.9
English Literature	36,913	2.7	3.8	15.8	30.9	44.4	1.1	1.3	6.6	15.7	26.6
Environmental Studies	877	14.5	1.3	13.5	24.4	41.2	12.7	3.1	5.4	16.3	19.0
Film Studies	5,765	15.4	6.4	26.0	40.9	50.5	11.8	1.9	11.3	27.1	41.1
French	7,589	8.6	33.5	62.1	72.9	79.2	2.5	20.3	47.4	66.6	75.2
Further mathematics	13,473	3.8	19.7	46.6	60.1	67.3	1.5	10.0	30.8	46.8	58.0
Geography	31,279	2.3	5.6	20.4	34.4	47.5	1.1	2.3	9.6	18.4	28.1
Geology	1,157	7.2	6.1	25.4	38.5	52.4	2.4	1.4	9.1	23.4	37.0
German	2,857	17.0	50.8	72.8	77.6	79.7	4.3	35.6	68.6	83.1	89.4
Greek (Modern)	186	62.4	24.2	37.6	37.6	37.6	44.1	38.7	45.7	45.7	55.9

Gujarati	25	40.0	60.0	60.0	60.0	60.0	40.0	16.0	60.0	60.0	60.0
History	47,044	1.7	3.3	12.7	24.2	35.2	0.9	1.1	5.4	11.9	19.3
Italian	799	33.3	42.6	60.7	64.7	66.7	20.5	43.4	63.1	74.7	77.5
Japanese	195	50.8	35.4	41.5	49.2	49.2	29.2	54.9	63.1	63.1	70.8
Latin	1,103	11.6	47.6	74.3	83.1	83.1	3.0	33.5	68.1	79.3	91.7
Law	11,092	4.0	3.5	14.2	24.2	31.3	2.9	1.9	6.5	13.6	21.7
Mathematics	83,898	1.2	1.5	5.5	10.4	16.0	0.9	0.6	2.5	5.5	8.8
Media Studies	14,004	5.1	4.5	21.2	41.4	55.0	3.7	1.4	8.9	21.7	35.2
Modern Hebrew	52	23.1	32.7	76.9	76.9	76.9	15.4	25.0	53.8	84.6	84.6
Music	5,056	9.4	37.8	59.2	70.6	76.9	3.0	24.8	52.4	65.4	72.1
Panjabi	182	27.5	22.5	35.7	44.0	44.0	19.2	24.2	44.0	44.0	52.2
Persian	205	42.4	31.2	33.2	43.4	43.4	31.2	33.2	44.4	44.4	44.4
Philosophy	2,768	12.4	8.2	31.0	46.7	56.7	11.6	5.4	19.8	33.6	48.1
Physics	36,025	2.3	5.7	20.5	34.0	46.2	1.2	2.6	9.9	19.9	29.2
Polish	1,004	36.6	41.4	49.4	52.7	56.1	17.1	42.2	63.3	68.4	72.1
Politics	18,091	6.9	5.6	21.4	42.2	55.7	4.1	2.5	9.8	23.0	34.9
Portuguese	494	39.1	36.8	53.2	57.5	60.9	24.7	41.7	61.1	69.4	75.3

Psychology	61,882	1.8	1.4	6.7	14.4	24.3	1.3	0.5	2.7	6.6	11.5
Religious Studies	16,220	4.4	8.6	34.3	53.6	68.8	1.0	3.3	15.0	32.3	47.9
Russian	695	25.0	44.2	65.9	67.5	72.7	13.1	37.8	59.6	73.7	75.3
Sociology	35,610	3.5	2.7	10.7	22.7	36.1	2.1	1.0	4.4	10.5	18.4
Spanish	7,786	11.8	28.3	59.9	70.5	75.2	2.6	18.1	45.6	66.2	76.3
Sport & PE	9,832	6.3	18.5	51.6	68.9	73.6	2.2	7.9	31.7	53.9	68.4
Statistics	669	24.4	4.0	10.9	20.9	27.5	21.5	1.6	6.7	9.7	25.3
Turkish	581	34.3	35.1	45.6	54.2	60.9	20.0	28.9	43.4	59.2	63.2
Urdu	401	14.5	11.0	44.1	50.1	50.1	8.0	13.7	31.7	49.9	56.6

Annex J: Modelled leniency based on 2019 data

Table J.1. Potential leniency for GCSE subjects due to small centres and those without historical data

		$n_{\text{thresh}} = 1, n_{\text{small}} = 10$			$n_{\text{thresh}} = 5, n_{\text{small}} = 15$			$n_{\text{thresh}} = 10, n_{\text{small}} = 20$		
Subject	**Number of candidates**	**Grade 7**	**Grade 4**	**Grade 1**	**Grade 7**	**Grade 4**	**Grade 1**	**Grade 7**	**Grade 4**	**Grade 1**
Art & Design	182,216	0.1	0.1	0.0	0.1	0.2	0.0	0.2	0.3	0.0
Biology	165,072	0.2	0.1	0.0	0.2	0.1	0.0	0.3	0.2	0.0
Chemistry	159,038	0.2	0.1	0.0	0.2	0.1	0.0	0.3	0.1	0.0
Citizenship Studies	19,644	1.0	1.0	0.2	1.1	1.2	0.2	1.2	1.3	0.2
Classical Greek	1,143	0.6	0.3	0.0	0.9	0.5	0.0	1.1	0.6	0.0
Computing	77,419	0.4	0.4	0.1	0.5	0.6	0.1	0.7	0.8	0.2
Dance	9,274	1.0	1.3	0.1	1.8	2.0	0.1	2.9	3.2	0.2
Drama	57,560	0.3	0.4	0.0	0.6	0.7	0.0	1.1	1.2	0.1
English Language	706,620	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.0
English Literature	546,638	0.1	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.0
Food Prep and Nutrition	44,921	0.3	0.6	0.0	0.5	0.8	0.0	0.9	1.4	0.1
French	121,756	0.2	0.3	0.0	0.2	0.4	0.0	0.4	0.6	0.1
Geography	250,986	0.0	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.0
German	40,655	0.3	0.4	0.0	0.4	0.5	0.1	0.7	0.7	0.1
History	260,468	0.1	0.1	0.0	0.1	0.1	0.0	0.1	0.1	0.0
Latin	9,355	0.4	0.3	0.1	0.7	0.4	0.1	1.2	0.4	0.1
Mathematics	719,874	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0
Music	34,049	0.5	0.6	0.0	1.3	1.3	0.1	2.4	2.5	0.2
Physical Education	79,942	0.3	0.4	0.0	0.4	0.5	0.0	0.6	0.7	0.0
Physics	157,787	0.2	0.1	0.0	0.2	0.1	0.0	0.3	0.1	0.0
Religious Studies	249,325	0.2	0.2	0.0	0.2	0.2	0.0	0.3	0.3	0.0
Spanish	95,860	0.3	0.4	0.1	0.3	0.6	0.1	0.5	0.7	0.1

Table J.2. Potential leniency for A level subjects due to small centres and those without historical data

		$n_{\text{thresh}} = 1, n_{\text{small}} = 10$			$n_{\text{thresh}} = 5, n_{\text{small}} = 15$			$n_{\text{thresh}} = 10, n_{\text{small}} = 20$		
Subject	**Number of candidates**	**A***	**A**	**E**	**A***	**A**	**E**	**A***	**A**	**E**
Accounting	2215	0.3	0.5	0.4	0.6	0.9	0.7	0.8	1.4	0.8
Ancient History	681	1.9	4.6	0.3	2.6	5.5	0.3	3.1	6.4	0.3
Arabic	765	4.1	3.4	1.0	4.8	4.2	1.3	5.2	5.0	1.6
Art & Design	38597	0.2	0.4	0.0	0.4	0.9	0.1	0.8	1.6	0.1
Bengali	28	8.8	2.0	2.1	9.6	2.1	2.1	9.6	2.1	2.1
Biblical Hebrew	66	6.4	2.8	0.0	9.1	5.6	0.0	10.4	7.1	0.0
Biology	63580	0.1	0.1	0.1	0.1	0.2	0.2	0.3	0.4	0.3
Business Studies	30236	0.1	0.3	0.1	0.2	0.7	0.1	0.4	1.3	0.2
Chemistry	54868	0.1	0.2	0.1	0.2	0.4	0.2	0.5	0.7	0.3
Chinese	2086	1.5	2.6	0.0	2.8	4.7	0.1	4.5	7.2	0.1
Classical Civilisation	2893	1.3	2.3	0.1	2.4	3.8	0.1	3.4	5.5	0.1
Classical Greek	210	9.5	2.4	0.0	12.8	3.2	0.0	13.4	3.4	0.0
Computing	10435	0.8	1.4	0.5	1.6	2.5	0.8	2.4	3.6	1.1
D&T: Product Design	9214	0.5	1.1	0.2	1.2	2.5	0.4	2.0	4.1	0.5
Dance	1016	1.3	3.0	0.1	2.6	5.6	0.1	3.5	7.4	0.2
Drama & Theatre Studies	8970	0.6	1.4	0.0	1.4	3.4	0.1	2.3	5.4	0.1
Dutch	33	4.8	8.6	1.6	5.5	9.1	1.8	5.5	9.1	1.8
Economics	29740	0.1	0.2	0.1	0.3	0.5	0.1	0.6	1.1	0.2
Electronics	588	0.4	0.4	0.0	1.1	1.1	0.0	2.0	1.8	0.2
English Language	13560	0.1	0.4	0.0	0.3	0.9	0.0	0.6	1.9	0.1
English Lang & Lit	7425	0.2	0.7	0.0	0.4	1.3	0.0	0.7	2.3	0.1
English Literature	36913	0.1	0.2	0.0	0.3	0.6	0.0	0.6	1.3	0.1
Environmental Studies	877	0.2	0.5	0.9	0.2	0.6	0.9	0.3	0.9	1.0
Film Studies	5765	0.5	1.6	0.0	0.7	2.5	0.1	1.1	3.9	0.1

French	7589	1.7	2.2	0.1	3.5	4.4	0.2	5.0	6.2	0.3
Further maths	13473	1.1	0.9	0.2	2.5	1.9	0.3	3.9	3.0	0.4
Geography	31279	0.2	0.3	0.0	0.4	0.8	0.1	0.8	1.5	0.1
Geology	1157	0.3	0.3	0.0	0.7	0.7	0.1	1.5	1.6	0.2
German	2857	3.5	2.9	0.1	5.9	5.0	0.1	7.3	6.2	0.1
Greek (Modern)	186	9.3	5.6	0.3	9.9	6.2	0.3	10.2	6.8	0.3
Gujarati	25	3.2	6.6	1.6	5.3	8.1	2.2	6.0	8.4	2.4
History	47044	0.1	0.2	0.0	0.2	0.5	0.0	0.5	1.1	0.1
Italian	799	6.7	2.8	0.2	8.7	3.8	0.2	9.8	4.4	0.2
Japanese	195	8.3	3.0	0.7	9.2	3.5	1.0	9.2	3.5	1.1
Latin	1103	4.7	2.7	0.2	7.8	4.4	0.2	9.9	5.3	0.2
Law	11092	0.1	0.3	0.2	0.2	0.6	0.2	0.4	1.0	0.3
Mathematics	83898	0.1	0.1	0.1	0.2	0.2	0.1	0.3	0.3	0.2
Media Studies	14004	0.1	0.6	0.0	0.3	1.2	0.0	0.6	2.4	0.1
Modern Hebrew	52	2.5	2.5	0.2	3.6	4.6	0.6	4.7	5.6	1.0
Music	5056	1.3	2.3	0.2	2.3	4.1	0.3	2.9	5.3	0.4
Panjabi	182	2.9	2.5	0.5	3.9	3.2	0.5	4.4	3.4	0.5
Persian	205	5.7	2.2	1.2	6.5	2.5	1.3	6.7	2.5	1.3
Philosophy	2768	0.6	1.2	0.3	1.1	2.0	0.4	1.8	3.1	0.5
Physics	36025	0.2	0.3	0.2	0.4	0.6	0.4	0.8	1.2	0.6
Polish	1004	8.0	3.2	0.2	10.6	4.5	0.3	11.6	5.1	0.3
Politics	18091	0.3	0.6	0.1	0.6	1.2	0.1	1.2	2.2	0.2
Portuguese	494	4.7	5.1	1.1	5.7	6.3	1.3	6.1	7.1	1.5
Psychology	61882	0.1	0.2	0.0	0.1	0.3	0.1	0.2	0.5	0.1
Religious Studies	16220	0.3	0.5	0.1	0.8	1.4	0.1	1.6	2.7	0.2
Russian	695	9.4	0.9	0.0	12.6	1.1	0.1	14.4	1.3	0.1
Sociology	35610	0.1	0.3	0.0	0.2	0.5	0.1	0.4	1.0	0.1
Spanish	7786	1.6	2.1	0.1	3.3	4.1	0.2	4.9	5.8	0.2
Sport & PE	9832	0.4	0.9	0.2	1.0	2.3	0.4	1.7	3.9	0.6
Statistics	669	0.5	1.3	0.2	0.5	1.5	0.4	0.5	1.8	0.6

Turkish	581	3.8	3.9	0.4	4.6	5.3	0.7	5.2	6.7	0.8
Urdu	401	2.3	1.1	0.1	4.2	2.3	0.1	5.9	3.3	0.1

Annex K: Appropriateness of statistical adjustments for small entries

Mike Cresswell Professional Services

Standardisation of GQ grades in Summer 2020

A short technical note on handling small centre entries

The agreed process for standardising the Centre-assessed grades (CAGs) being collected in Summer 2020 because of the cancellation of examinations due to the Covid-19 epidemic, involves finding criteria for identifying centre entries that are so small that no standardisation can justifiably be applied. However it would be undesirable to operate a sharp boundary, based upon entry size, between centre entries where standardisation is not done at all and those where the CAGs are fully standardised to have precisely the same distribution as the grades predicted by the standardisation model (model-predicted grades, hereafter called MPGs). It has therefore been agreed to taper the application of the model for centres between n_{thresh}, the size of entry at and below which there is no warrant to standardise at all, and n_{small}, the size of entry at and above which full standardization operates.

This note records some technical aspects relevant to the identification of a suitable value for n_{thresh}.

The identification of n_{thresh} requires the identification of the size of centre for which there can be no warrant for a standardisation process. In this paper, the operationalisation of this question is taken to be: what size of entry is so small that it is impossible to have 95% confidence that a difference between the CAGs and MPGs is the result of bias in the CAGs, rather than the inevitable random variation between years in grade distribution which is to be expected with small numbers of candidates?

To answer this question, the first issue that needs to be addressed is the identification of an appropriate statistical test to apply to very small individual centre subject entries. For fewer than 10 candidates, it would be a heroic assumption that the underlying distribution of differences between CAGs and MPGs followed any specific distribution, especially since very high degrees of skew are likely for the cases where there is a big difference between CAGs and MPGs. Since tests for non-normality, like the Kolmogorov-Smirnov, are notoriously lacking in power, using that approach with very small numbers would risk assuming normality where it did not actually exist. It follows that it would be unwise to use parametric tests of significance.

Second, there is the issue of correlation. The CAGs and MPGs will be highly correlated within a centre. Not only do they relate to the same students but the standardisation approach guarantees a high correlation because it uses the centre rank order (which is, by definition, monotonically related to the CAGs) to assign MPGs to individual candidates. So any significance test used must be one suitable for testing differences between correlated data sets. That means that several well-known non-parametric significance tests (for example, the Mann-Whitney U test) would be inappropriate.

Of the standard significance tests that are left, the Wilcoxon Signed-Rank test is generally recommended in a context such as the present one, with sample points on at least ordinal scales. Figure 1 shows a portion of the Wilcoxon Signed-Rank test critical values table (two tailed because we are concerned with significant severity as well as leniency).

Figure 1 – Critical values for the Wilcoxon Signed-Rank test

	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
5	--	--	--	--	--	0	2
6	--	--	--	--	0	2	3
7	--	--	--	0	2	3	5
8	--	--	0	2	3	5	8
9	--	0	1	3	5	8	10
10	--	1	3	5	8	10	14
11	0	3	5	8	10	13	17
12	1	5	7	10	13	17	21
13	2	7	9	13	17	21	26
14	4	9	12	17	21	25	31
15	6	12	15	20	25	30	36
16	8	15	19	25	29	35	42
17	11	19	23	29	34	41	48
18	14	23	27	34	40	47	55
19	18	27	32	39	46	53	62
20	21	32	37	45	52	60	69
21	25	37	42	51	58	67	77
22	30	42	48	57	65	75	86
23	35	48	54	64	73	83	94
24	40	54	61	72	81	91	104
25	45	60	68	79	89	100	113
26	51	67	75	87	98	110	124
27	57	74	83	96	107	119	134

Figure 1 shows that with an N of 5 or fewer, there is no value low enough to produce significance at the 5% level or beyond. A Wilcoxon test result (T) of 0 would come from distributions in which the cumulative proportion of CAGs is either higher at EVERY grade or lower at EVERY grade than the cumulative proportion of MPGs. Even very extreme patterns of difference, such as wildly disjoint CAGs and MPGs (eg CAGs of A* A B B C and MPGs of D D E E U or even 5 grade 9s vs 5 grade Us!) only produce a Wilcoxon T of 0 because it cannot become negative. So the significance test of choice for these data simply isn't powerful enough for samples of 5 or fewer to decide that any difference between CAGs and MPGs is significant.

This provides good reason for setting n_{thresh} at a value of 5 because that is the highest number at which it is statistically impossible to be 95% certain that there is a significant difference between the CAGs and MPGs and therefore impossible to be 95% certain that the CAGs warrant adjustment.

The choice of a value of 5 for n_{thresh}, is therefore reasonable.

Mike Cresswell
Woking
June 2020

Annex L: Adjustment procedure in the presence of off-tier grades

For GCSE tiered subjects with at least 500 prior-attainment matched candidates (since subjects with smaller entries will not be adjusted to meet prediction), the procedure for adjusting cut-scores in the presence of off-tier grades is as follows:

i. Follow the standardisation process as described in Section 8.2

ii. Determine the number of prior-attainment matched students from the higher tier achieving each foundation tier (off-tier) grade, denoted as N'_{H2} and N'_{H1}

iii. Determine the number of prior-attainment matched students from the foundation tier achieving each higher tier (off-tier) grade, denoted as N'_{F6}, N'_{F7}, N'_{F8} and N'_{F9}

iv. Produce a national imputed mark distribution for matched candidates entered for the foundation tier in the subject

v. Produce a national imputed mark distribution for matched candidates entered for the higher tier in the subject

vi. From the mark distribution generated in iv., identify the candidate that falls N'_{H2} candidates below the current cut-score and set the revised grade 2 cut-score to match that student's imputed mark.

vii. Repeat step vi. for grade 1 using N'_{H1}

viii. From the mark distribution generated in v., identify the candidate that falls N'_{F6} candidates below the current cut-score and set the revised grade 6 cut-score to match that student's imputed mark.

ix. Repeat step viii. for grade 7 using N'_{F7}

x. Repeat step viii. for grade 8 using N'_{F8}

xi. Repeat step viii. for grade 9 using N'_{F9}

xii. Exam boards regrade students based on the revised cut-scores

Annex M: Standardisation at arithmetic grades

In a typical year, exam boards focus their attention on the *judgemental* grades[137] for the purposes of setting and maintaining standards. This enables sufficient statistical control and thorough scrutiny of students' work in a deliverable way to ensure that standards are maintained. A consequence of this approach is that the intermediate grade boundaries are set arithmetically – meaning they are, notionally, evenly spaced.

The situation in summer 2020 is different, given the absence of work to scrutinise or grade boundaries to set. Using any of the approaches to statistical standardisation that have been considered for use this summer, outcomes at every grade could be considered independently (and potentially adjusted independently) for each centre in each subject. One advantage is that this approach provides control at all points across the grade scale. This advantage might also, however, be considered a disadvantage given the risks of spurious precision. For example, a centre whose CAGs are broadly in line with their statistical prediction may receive small upward and downward adjustments at each grade with the statistical model being applied at all grades. This has the potential to remove legitimate, but undetectable, features of the centre's grade distribution.

It could be argued that ensuring these fine adjustments are made only at the judgemental grades provides sufficient protection against overall inflation/deflation. The arithmetic grades could then be adjusted "smoothly" and consistently across the grade scale based on the interpolation of the adjustments required at the judgemental grades.

In the simplest terms, a centre entering a GCSE subject and receiving an adjustment to their CAGs of -12% at grade 7 and -3% at grade 4 would lead to grades 5 and 6 being adjusted by -6% and -9%, respectively.

This section presents the results of applying different approaches to realising a smooth adjustment at the arithmetic grades based on the statistical adjustments required at key grades[138].

The simple approach

The simplest approach to realising a smooth adjustment at the arithmetic grades would be to apply the formula below:

$$P_{kj}^{*} = \max\left(P_{aj}, \min\left(P_{bj}, f_{kj} + \left(P_{bj} - f_{bj}\right) + \left((P_{aj} - f_{aj}) - (P_{bj} - f_{bj})\right)\frac{(k-b)}{(a-b)}\right)\right)$$

(eq.1)

137 For the purposes of awarding, exam boards routinely focus both the evaluation of statistical evidence and the scrutiny of candidates' work at a few grades termed judgemental grades. At GCSE, these are Grades 7, 4 and 1. At AS and A level, these are grades A and E. Statistical monitoring is also performed at Grade 9 and A*.

138 The term "key grades" is used here to reflect the potential flexibility of applying the statistical approaches at grades other than those conventionally used as judgemental grades.

where:

- k is the arithmetic grade to be adjusted[139]
- a is the lowest key grade above the arithmetic grade
- b is the highest key grade below the arithmetic grade
- P^*_{kj} is the revised predicted percentage of students achieving grade k or above in centre j
- P_{aj} and P_{bj} are the percentages of students achieving grades a and above and b and above, respectively, after statistical adjustment
- f_{kj}, f_{aj} and f_{bj} are the percentages of students estimated to achieve each grade and above, based on the submitted CAGs

Note that the max and min functions are necessary to avoid any instances of disordered outcomes.

Data from 2019 arising from the testing of the DCP approach are used as a basis for the analysis presented here. In the absence of authentic CAG data at the time of development, and to understand the impact of CAGs with different profiles, it was necessary to simulate these data.

Given the likely generosity in the CAGs relative to the outcomes that would have been actually achieved had exams taken place, the following approach to simulating CAG profiles has been applied initially:

$$\hat{f}_{kj} = \begin{cases} q_{kj} + \delta_j q_{k-1j}, & \text{if } k = M \\ q_{kj} - \delta_j q_{kj} + \delta_j q_{k-1j}, & \text{if } 0 < k < M \\ q_{kj} - \delta_j q_{kj}, & \text{if } k = 0 \end{cases}$$

where M is the highest grade for the qualification, δ_j is the inflation factor for centre j expressed as a proportion, q_{kj} is the percentage of students at centre j predicted to achieve grade k exactly, and $\hat{f}_{kj}$ is the percentage of students at the centre with CAGs at grade k exactly.

The effect of this approach is to promote $100 \times \delta_j$ percent of students from each grade to the grade above.

For the purposes of simulation here, a value of 0.3 have been applied to all centres. This approximation does not appear unreasonable based on the levels of leniency quoted in the literature.

139 For adjustment purposes, the grades are converted to numerical values as follows: A level – A* = 6, A = 5…E = 1 and U = 0; GCSE – Grade 9 = 9, Grade 8 = 8…Grade 1 = 1, U = 0.

Provided in Figure M.1., for each A level subject, is the variation of subject-level outcome relative to prediction at the arithmetic grades, plotted against entry size having applied equation 1. This analysis has the key grades set as A*, A and E (meaning the predictions are met exactly at these grades).

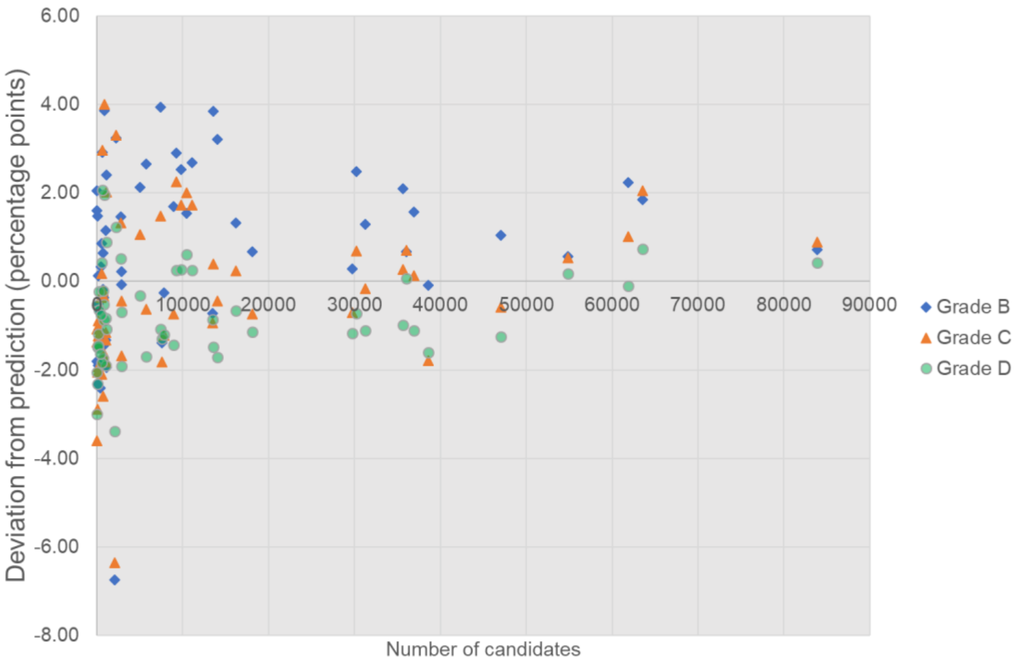


Figure M.1. Subject-level leniency/severity v entry size at the arithmetic grades for each A level subject based on a generosity rate of 30% and full statistical adjustments applied at grades A, A and E*

While there is seemingly some relationship between the variation in outcomes from prediction with entry size, this relationship appears relatively weak. More informative in terms of understanding the cause of the leniency/severity at the arithmetic grades is the plot shown in Figure M.2. This shows the relationship between the deviation from prediction plotted against the predicted subject-level outcome at that grade.

This figure shows that there is a relationship between the leniency/severity of the outcomes at the arithmetic grades and the predicted outcome for the subject, with the likelihood of severe outcomes increasing as the predicted outcome increases. This tendency is shown within and across grades.

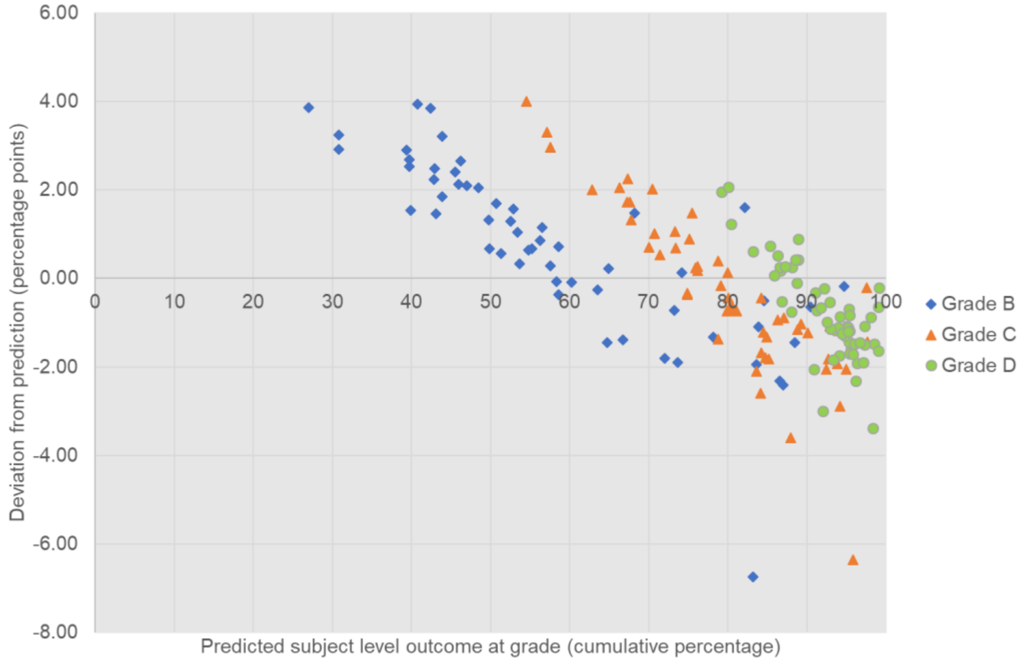


Figure M.2. Subject-level leniency/severity v predicted outcome at the arithmetic grades for each A level subject based on a generosity rate of 30% and full statistical adjustments applied at grades A, A and E*

The equivalent plot for GCSE subjects[140], with the key grades set to grades 9, 7, 4 and 1, is provided in Figure M.3. This figure shows varying levels of leniency/severity at different grades, with a less clear relationship with predicted outcome. One of the most notable features of this plot is the tendency for outcomes at grades 2 and 3 to be severe of prediction having the potential to systematically disadvantage students at these grades. This issue will be returned to below.

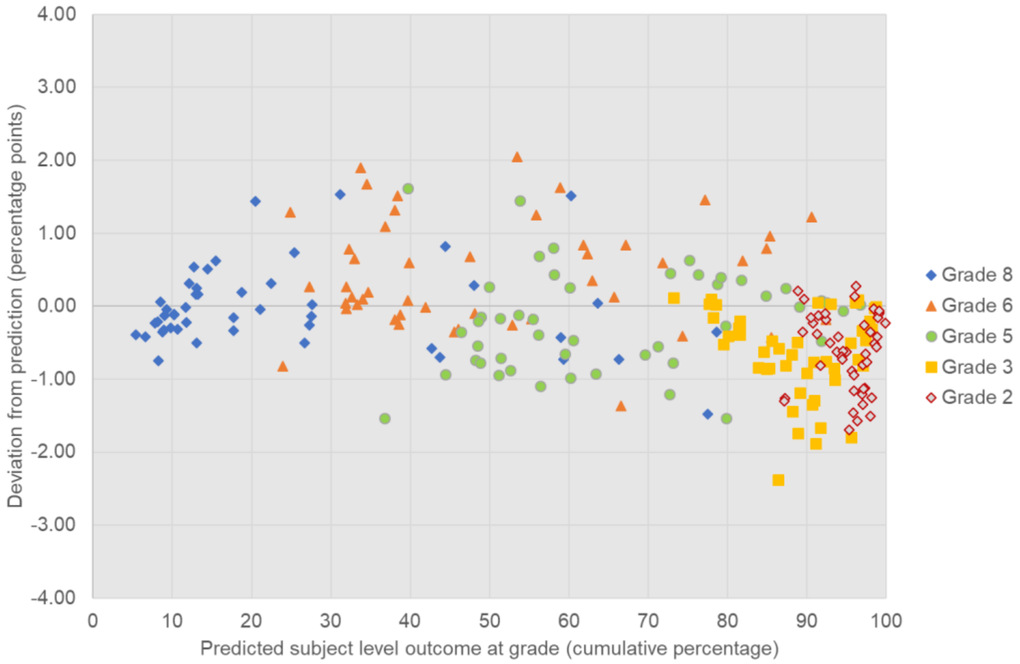


Figure M.3. Subject-level leniency/severity v predicted outcome at the arithmetic grades for each GCSE subject based on a generosity rate of 30% and full statistical adjustments applied at grades 9, 7, 4 and 1

Having explored the data under-pinning Figures M.2. and M.3. and data from similar simulations, there appear to be 4 key factors which influence the tendency for leniency/severity at the arithmetic grades:

i. **The concentration effect.** The differential between the number of students “incoming” to the grade from the grade below and the number of students “outgoing” to the grade above.

ii. **The lever effect.** The impact of a differential concentration at the key grades either side of the arithmetic grade.

iii. **The profile effect.** The effect of a varying profile of generosity across the grade distribution (i.e. δ_j not being constant across grades for a centre such that δ_j becomes δ_{kj}). (Note that this effect is not present in the simulated results presented above).

140 Note that for the purposes of the work presented here, the GCSEs considered were those subjects in phase 1 and phase 2 of qualifications reform due to the availability of data for reformed specifications in 2019.

iv. **The centre-level distribution effect.** The specific grade distribution of centres within a subject impacts the opportunity for and profile of generosity. As can be seen from Figure 2, there are ceiling effects as the cumulative percentage outcomes for a subject approach 100%. This effect also occurs at the individual centre-level affecting different centres in different ways.

Effects i., ii., and 0., are present in the simulation results presented above and interact heavily to dictate the levels of leniency/severity observed.

The concentration effect

This effect arises from the differential concentration of students across the grade scale. This means that, in different areas of the distribution, there are more or fewer students available for promotion to a higher grade to represent generosity. It could be argued that this effect is an artefact of the simulation, due to a uniform proportion of students being moved from one grade to the next to simulate generosity. In practice, other 'real-life' factors, such as the stakes of students achieving the different grades (for example, at grade 4 in GCSE English language and GCSE mathematics) will be overlaid on this uniform profile of generosity (leading to the profile effect highlighted above). This profile effect will be considered separately. To an extent it is true that the concentration effect is an artefact of the simulation process, however, this underlying variation in concentration of students will exist in practice and have an impact on the absolute levels of generosity that can occur across the grade range. As this will interact with the other effects discussed here, it is worthy of consideration.

In simple terms, where the source of students to feed generosity at a grade is less plentiful than the number of students to be passed on from that grade, there will be a deficit in students. For example, at GCSE, more students will be out-going from grade 1 to grade 2 than will be incoming to grade 1 from being ungraded. In other areas of the distribution, such as grade 7, the opposite is true. Here, more students will be incoming from grade 6 than would be outgoing to grade 8. This effect is illustrated in Figure M.4. which shows the net change in students at each grade aggregated across all GCSEs with a generosity factor of 0.3. It is noted that this visual representation of the data is not technically sound due to the categorical rather continuous nature of the data, however, this is helpful for visualisation of the effects described.

Figure M.4. Net changes in the percentage of students at each grade based on the differential concentration of students across the grade distribution with δ = 0.3

Overlaid on this plot, in Figure M.5., are the values of leniency/severity aggregated across all GCSEs. This shows that, following the adjustment process, the differences from prediction at grades 9, 7, 4 and 1 reduce to zero. However, the outcomes relative to prediction at the arithmetic grades vary depending on the concentration of students at the key grades that surround it. For example, the concentration effect leads to a larger surplus at grades 9 and 7 than is the case at grade 8. Once the standardisation is applied at the key grades, this differential effect leaves grade 8 severe of prediction.

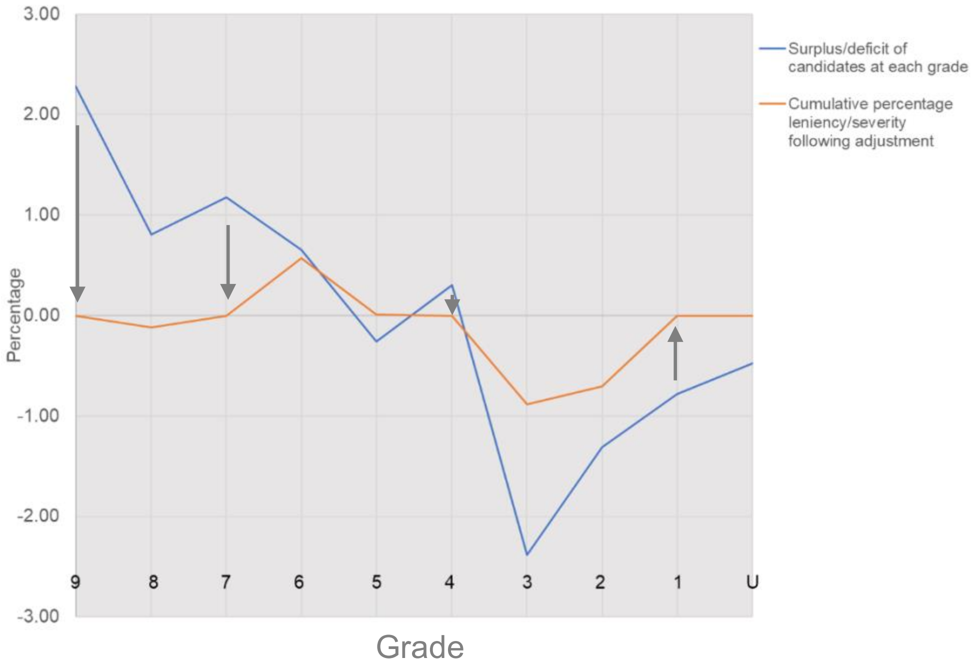


Figure M.5. Net change in students across the grade range following application of a uniform generosity factor and the impact of this distribution at the arithmetic grades following standardisation

This is not the only effect at play, however, and the relative leniency/severity cannot be simply described by this relationship. Another factor having an impact is the *lever effect* considered below.

The lever effect

This level effect is demonstrated here through the use of an artificially extreme case. Shown in Figure M.6. is the leniency that results from applying the generosity factor of 0.3 only at grade B for all A level subjects. This means that 30% of students predicted to achieve a grade C are moved up to a grade B in the CAGs. The key grades are initially defined as A*, A and E in line with typical practice. This plot shows that, as there is no generosity at grades A and E, no generosity is deemed to exist in the CAGs at the other grades. This means the full effect of the generosity at grade B propagates through to the final outcomes.

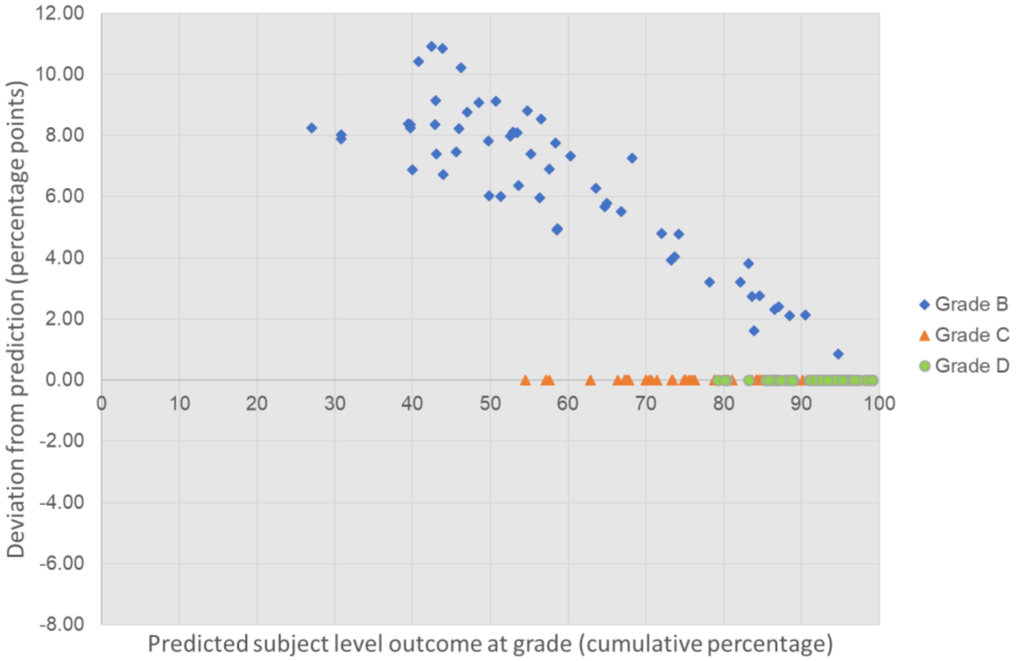


Figure M.6. The generosity when applied only at grade B with key grades set to A, A and E*

Adding grade B as a key grade, however, demonstrates the lever effect and how it impacts on the other arithmetic grades. This is shown in Figure M.7. This shows the dramatic impact on grades C and D due to the large differential in adjustment applied at the key grades that surround them (i.e. a large adjustment at grade B and no adjustment at grade E) with legitimately graded students at grades C and D being forced lower than would otherwise be the case.

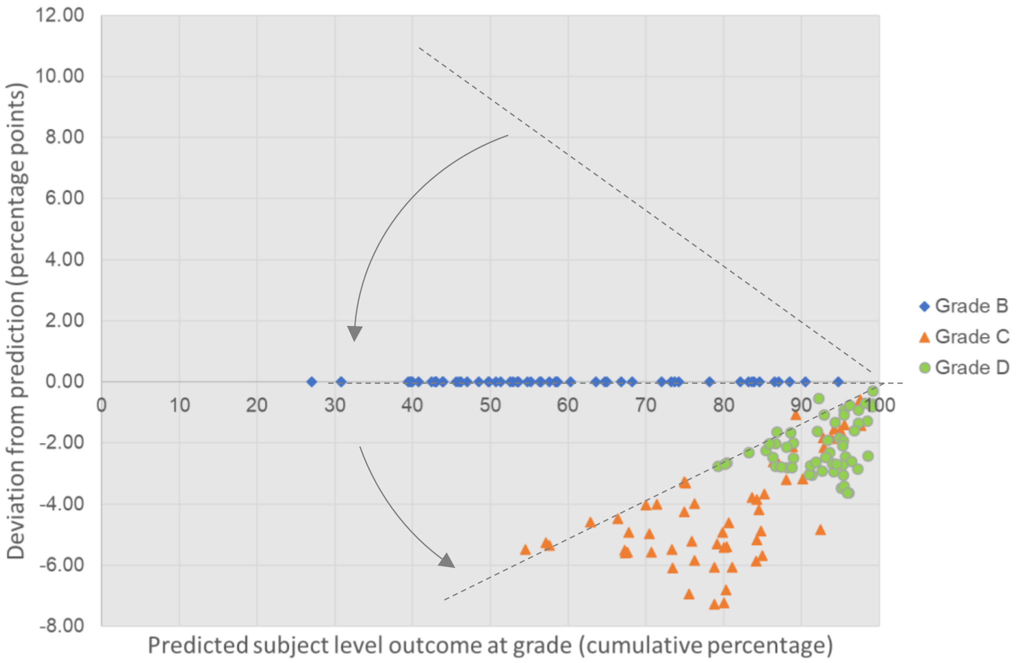


Figure M.7. Demonstration of the lever effect resulting from differential adjustments being applied at key grades

For ease of explanation, this effect has been demonstrated artificially here with a stark non-uniform profile of generosity applied across the grade scale and by changing the key grades to remove other effects related to the shape of the distribution. However, this effect will occur organically in the plots of leniency/severity presented above due to the variations in student concentration shown in Figure M.5.

The profile effect

All of the effects shown here have been based on a generosity rate of 0.3 applied across the grade scale. As described above, it is unlikely that, in practice, the same level of generosity will exist at all grades – partly due to natural variations and partly due to the different external drivers. Given the complex interactions that take place in the data underlying the figures plotted here, it is difficulty to isolate a single case of non-uniform generosity that generalises in a meaningful way.

A useful case study, however, is illustrated in Figure . This shows the difference in leniency/severity between 2 cases applied across GCSE subjects:

1) applying a 0.3 generosity factor at all grades
2) applying a 0.3 generosity factor at all grades with the exception of grade 4 where a factor of 0.5 is applied

Despite the absence of a deviation from prediction at grade 4 (due to it being a key grade) the impact on the arithmetic grades either side is significant due to the combination of effects explored above.

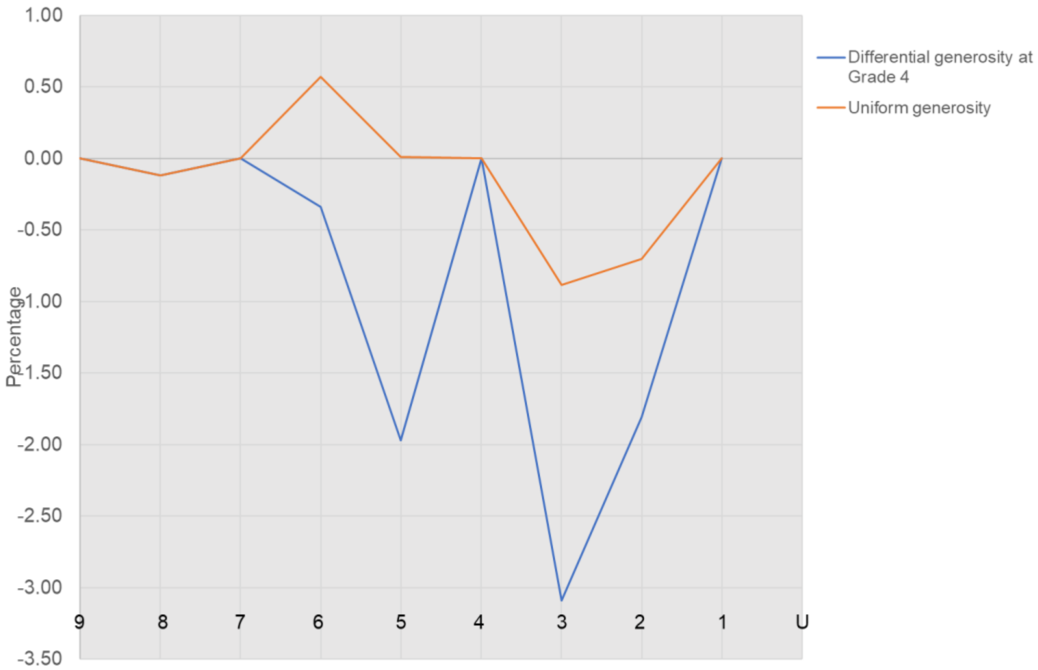


Figure M.8. Effect of applying a differential level of generosity at grade 4 (50%) with uniform generosity at other grades (30%)

Focusing on the lower part of the grade distribution, this demonstrates the potential unfairness applying the simple approach defined in equation 1 could have. Here, students with CAGs at grades 2 and 3 may be severely dealt with relative to prediction due to the high level of adjustment required at grade 4.

Using a combination of real data and simulated effects it is challenging to isolate and demonstrate the magnitude and direction of these different effects in a controlled way. To do this thoroughly and precisely would require the analysis of a fully simulated and fully manipulatable data set. However, given the concerns outlined above, this detailed analysis is not currently necessary beyond identifying the need to explore how these effects might be mitigated while still being able to realise the initial aim of applying a smooth adjustment across the grade scale.

The tapered approach

Outlined above were the technical challenges of applying a simple approach to adjusting the arithmetic grades based on the adjustments required at the surrounding key grades. In addition to these technical issues, there are conceptual challenges too. An important aim applying adjustments smoothly between key grades is to better reflect the distribution of students within a centre based on the CAGs they submitted. It can be argued, however, that the plausibility of the CAGs reduces the further they are from the statistical prediction. It is, therefore, a logical extension that as the plausibility of the outcomes reduces, so too should the confidence in preserving the relative (albeit scaled) grade distribution for the centre.

The approach considered here seeks to simultaneously mitigate the technical issues identified above and this conceptual challenge of potentially over-interpreting the CAGs as their plausibility reduces. To do this, a tapered approach to preserving the centre-level grade distribution was considered. This would mean the level of smoothing at arithmetic grades would be dependent on the distance of the centres' CAGs from the statistical predictions.

The effect of this is best explained through consideration of extreme cases. In instances where there is no net generosity or severity in the CAGs, the relative distribution suggested by the centre at the arithmetic grades would be retained. In instances where the CAGs represent a highly generous or highly severe estimation of outcomes relative to the statistics, the full statistical adjustments would be applied at each grade. Between these two extremes, the balance between the distribution suggested by the CAGs and the statistical evidence would be tapered depending on the net generosity/severity in the CAGs.

To follow this approach, it was necessary to determine a summary statistic to characterise the generosity or severity for each centre in the subject. The statistic used was defined by a summation across the grade range of the number of students achieving a higher grade based on the CAGs than is predicted by the model:

$$\rho_{\text{gen}} = \frac{\sum_{k \in G}\left(f_{kj} - P_{kj}\right)}{\sum_{k \in G}\left(P_{k-1j} - P_{kj}\right)}$$

where ρ_{gen} is the generosity coefficient for the centre in the subject, $k-1$ is the grade lower than k and G is the grade set across which the coefficient is calculated. There were, however, two potential approaches to defining the grade set over which calculation takes place:

Option 1) Across the key grades only

Option 2) Across all grades

Option 1, in many ways, is in-keeping with the rationale for seeking to apply the statistical adjustment at the key grades only and allow greater freedom where it has been shown that the overall leniency/severity is low. The downside, however, is that this approach puts particular significance on certain areas of the grade distribution which interacts with the grade distribution for the individual centres. This could potentially advantage/disadvantage different centres based solely on their profile of ability. For example, considering A level, were the conventional key grades used in the calculation of ρ_{gen}, the coefficient would be based solely on the relationship between the CAGs and statistical prediction at grades A* and A for the majority of centres as most centres have very few (if any) students predicted to be ungraded to provide a source of generosity at grade E.

In contrast, **Option 2)** enables information to be drawn from across the grade distribution and has the effect of determining the difference between the mean grade defined by the CAGs and that predicted by the statistical model.

Option 2 was pursued given the added stability in the calculation it provides and its reduced dependency on the grade profile of individual centres. This simplifies the calculation of ρ_{gen} to:

$$\rho_{\text{gen}} = \sum_{k=1}^{M}\left(f_{kj} - P_{kj}\right)$$

Once ρ_{gen} is calculated, it can be used to define the weight to be put on the CAGs at the arithmetic grades compared to the statistical evidence. This is achieved as:

$$\hat{P}_{kj} = \begin{cases} \left(1-\frac{|\rho_{\text{gen}}|}{\rho_{\text{thresh}}}\right)P^{*}_{kj} + \left(\frac{|\rho_{\text{gen}}|}{\rho_{\text{thresh}}}\right)P_{kj}, & \text{if } |\rho_{\text{gen}}| \le \rho_{\text{thresh}} \\ P_{kj}, & \text{if } |\rho_{\text{gen}}| > \rho_{\text{thresh}} \end{cases}$$

where $\hat{P}_{kj}$ is the final predicted cumulative percentage at arithmetic grade, k and ρ_{thresh} is the highest value of ρ_{gen} at which the CAGs influence the centre-level prediction.

To model this approach effectively, it is necessary to vary the generosity/severity across centres. To do this, values of the generosity factors (δ_j) were randomly sampled for each centre in each subject from $\mathcal{N}(0.167, 0.167)$. These values were selected to reflect average levels of leniency approximated from the research literature referenced in Section 3. The value of ρ_{thresh} was set to $^1/_3$ such that this average level of potential over-prediction occurs at the point where the balance between the CAGs and the full statistical adjustment is equal. The results, with and without tapering, for all A level subjects at grades B, C and D are shown in Figure M.9.

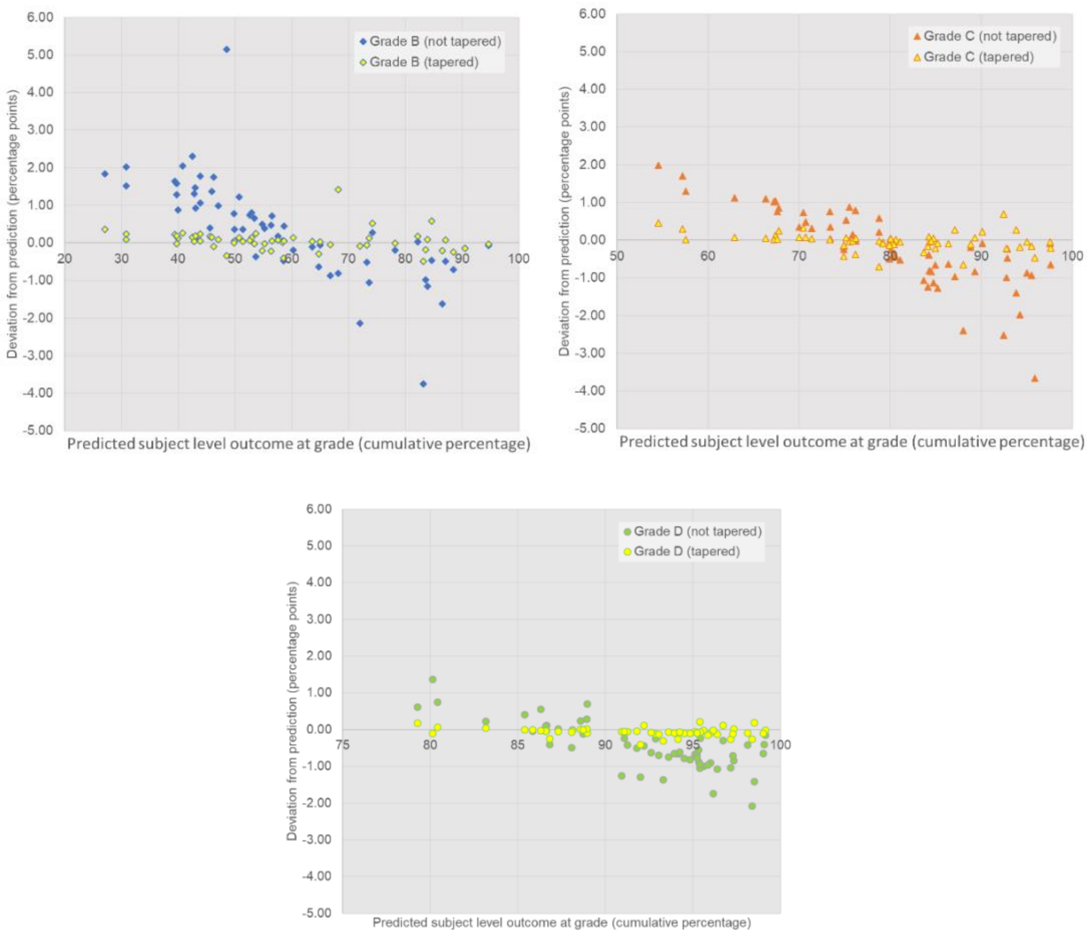


Figure M.9. The impact on leniency/severity of applying the tapered approach compared to the simple approach for applying adjustments at the arithmetic grades

These results show a significant reduction in the undesirable deviation from prediction when applying the tapered approach by suppressing the unpredictable effects as centres' CAGs deviate increasingly far from prediction. These results are summarised in Figures M.10. and M.11., having been aggregated across all A level and GCSE subjects, respectively.

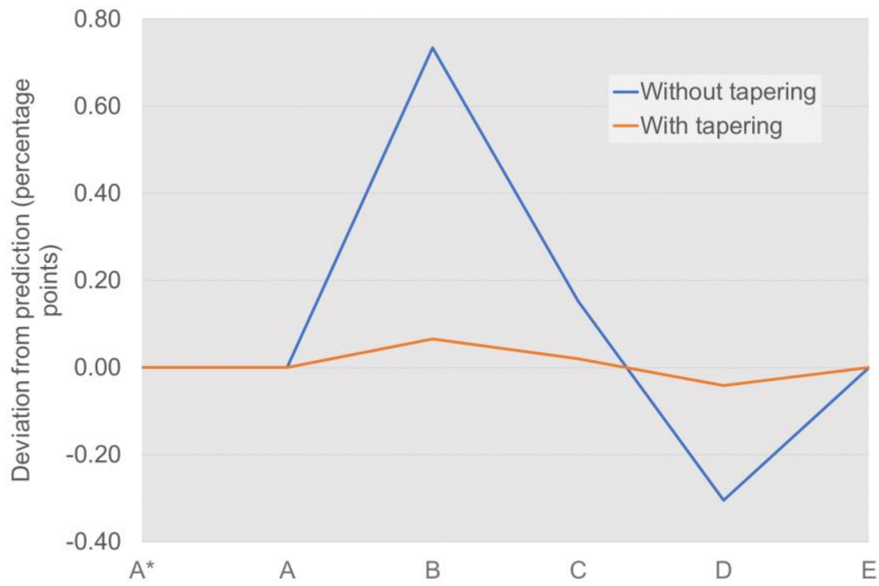


Figure M.10. The reduction of overall deviation from prediction at the arithmetic grades, aggregated across all A levels, caused by taking the tapered approach to adjustment

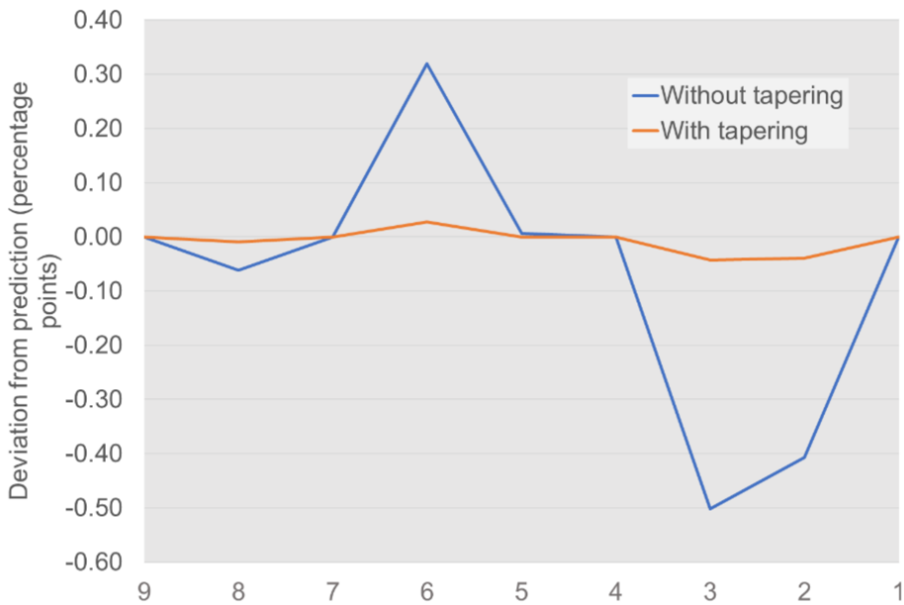


Figure M.11. The reduction of overall deviation from prediction at the arithmetic grades, aggregated across all GCSEs, caused by taking the tapered approach to adjustment

The tapered approach to adjustment, therefore, emerged as a promising approach to achieving a smooth adjustment for a centre within each subject. However, following completion of the testing described here, provisional CAG data became available to better understand the impact of applying the approach using real data. Shown below in Figure M.12. are the results of applying both the simple and taped approaches to adjusting arithmetic grades for GCSE English language based on real CAGs following collection for approximately 75% of the entry.

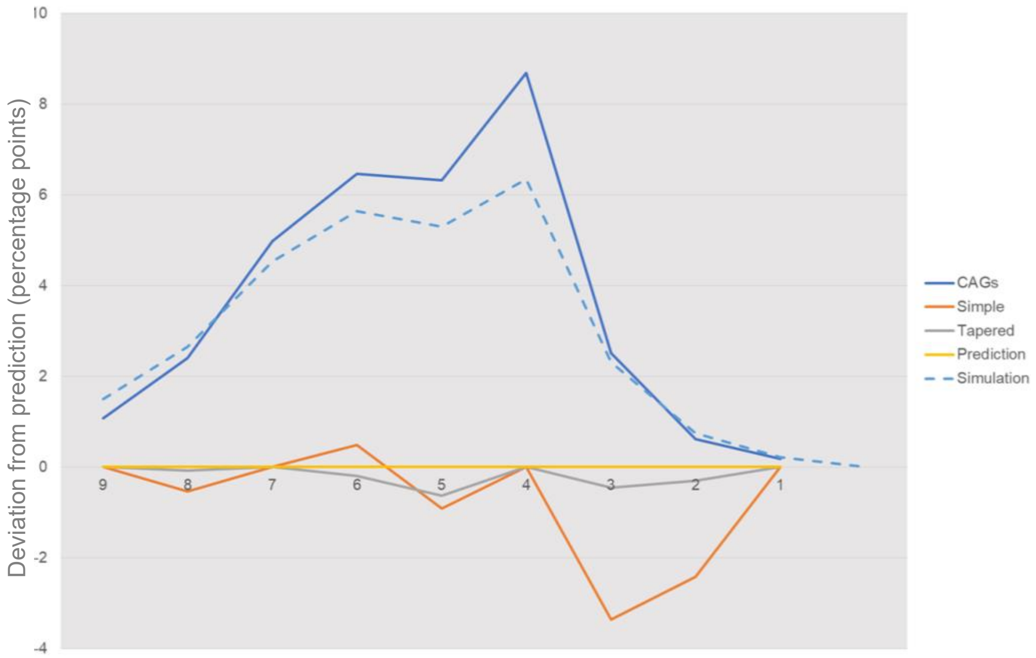


Figure M.12. Impact at different grades of applying the simple and tapered adjustments at arithmetic grades for GCSE English language based on real CAG data

In addition to showing the outcomes relative to prediction for the two versions of the arithmetic grade adjustments are: the outcomes were the CAGs to have been accepted (solid blue line) and the simulated levels of generosity (dashed blue line). It can be seen from the difference between these two blue lines that the profile of the generosity in the CAGs was non-uniform compared to the uniform approach applied in the simulations. This showed that, due to this profile, irrespective of the use of the tapered approach, this risked inadvertent severity at the grades surrounding grade 4. Importantly, if using the tapered approach, it should be noted that this effect would only impact on those within the tapered region ($|\rho_{\text{gen}}| < {}^{1}/_{3}$) and, therefore, those whose CAGs were nearest to the statistical prediction. Those with far more generous CAGs would not be disadvantaged as they would receive the statistical adjustment at each grade (indicated by the yellow line in Figure M.12.).

Due to the potential disadvantage introduced by this approach where there is a systematic generosity in a centres' CAGs at different points in the grade scale, and the unpredictable nature of the effects discussed here, it was decided that the statistical adjustments should be applied at all grades using the same approach, rather than applying the alternative solutions investigated here.

Annex N: Centres with small cohorts by subject – A level

Table N.1. Percentage of centres with small cohorts by A level subject

Subject	Total number of centres	% of total centres		
		Small cohorts	Tapered adjustment	Statistical adjustment
Accounting	160	31.9	28.1	40.0
Ancient History	65	41.5	35.4	23.1
Arabic	120	85.0	12.5	2.5
Art & Design: 3D Studies	147	51.0	29.9	19.0
Art & Design: Art, Craft and Design	801	38.1	47.3	14.6
Art & Design: Critical & Contextual Studies	20	60.0	20.0	20.0
Art & Design: Fine Art	1,671	28.4	51.0	20.5
Art & Design: Graphics	397	32.5	34.3	33.2
Art & Design: Photography	1,235	28.5	45.9	25.6
Art & Design: Textiles	483	56.5	32.5	11.0
Biology	2,555	8.1	25.8	66.0
Business Studies	1,740	11.2	29.6	59.2
Chemistry	2,509	11.6	30.8	57.5
Chinese	423	59.6	24.1	16.3
Classical Civilisation	379	38.0	43.3	18.7
Classical Greek	88	89.8	9.1	1.1
Computing	1,404	35.3	46.7	18.0
D&T: Fashion and Textiles	153	77.8	21.6	0.7
D&T: Product Design	1,255	51.2	44.7	4.1
D&T: Design Engineering	47	55.3	40.4	4.3
Dance	214	50.0	41.1	8.9
Drama & Theatre Studies	1,309	30.0	54.2	15.7
Economics	1,657	13.3	26.5	60.2
Electronics	64	21.9	46.9	31.3
English Language	984	13.7	34.8	51.5
English Language & Literature	563	19.0	39.6	41.4
English Literature	2,451	11.8	33.3	54.9
Environmental Studies	60	30.0	35.0	35.0
Film Studies	511	18.8	45.8	35.4
French	1,519	51.2	39.8	9.1
Further Mathematics	1,924	38.4	43.8	17.8
Geography	2,100	16.2	38.1	45.7
Geology	105	14.3	50.5	35.2
German	791	69.7	28.2	2.1
History	2,470	11.3	33.1	55.7
History of Art	85	38.8	42.4	18.8
Italian	293	86.0	12.3	1.7
Japanese	54	92.6	5.6	1.9
Latin	308	72.4	25.0	2.6

Law	536	16.4	28.0	55.6
Mathematics	2,730	9.1	20.0	70.9
Media Studies	993	10.2	37.3	52.6
Music	1,059	68.9	27.9	3.2
Music Technology	240	51.7	37.9	10.4
Panjabi	18	66.7	27.8	5.6
Philosophy	240	28.8	44.2	27.1
Physical Education	1,290	28.1	52.8	19.1
Physics	2,432	16.6	37.7	45.7
Polish	214	89.3	10.7	0.0
Politics	1,341	19.6	37.7	42.7
Portuguese	78	92.3	5.1	2.6
Psychology	2,403	8.3	20.4	71.2
Religious Studies	1,487	17.8	49.0	33.3
Russian	253	77.1	19.0	4.0
Sociology	1,715	10.6	24.5	64.8
Spanish	1,434	48.5	40.2	11.3
Statistics	55	30.9	21.8	47.3
Turkish	85	81.2	10.6	8.2
Urdu	56	75.0	17.9	7.1

Annex O: Centres with small cohorts by subject - AS

Table O.1. Percentage of centres with small cohorts by AS subject

Subject	Total number of centres	% of total centres		
		Small cohorts	Tapered adjustment	Statistical adjustment
Accounting	37	51.4	27.0	21.6
Ancient History	10	40.0	10.0	50.0
Art & Design: 3D Studies	20	70.0	5.0	25.0
Art & Design: Art, Craft and Design	105	65.7	24.8	9.5
Art & Design: Fine Art	229	70.7	20.1	9.2
Art & Design: Graphics	49	65.3	16.3	18.4
Art & Design: Photography	215	60.5	27.9	11.6
Art & Design: Textiles	48	68.8	22.9	8.3
Biology	628	62.9	14.3	22.8
Business Studies	383	57.7	19.3	23.0
Chemistry	601	60.7	18.5	20.8
Chinese	83	88.0	8.4	3.6
Classical Civilisation	38	60.5	26.3	13.2
Computing	248	66.9	21.4	11.7
D&T: Product Design	82	63.4	26.8	9.8
Drama & Theatre Studies	80	57.5	27.5	15.0
Economics	313	57.8	18.2	24.0
Electronics	26	57.7	23.1	19.2
English Language	165	51.5	20.6	27.9
English Language & Literature	80	43.8	27.5	28.8
English Literature	272	46.7	24.6	28.7
Environmental Studies	24	62.5	16.7	20.8
Film Studies	74	39.2	24.3	36.5
French	238	73.5	20.6	5.9
Further Mathematics	816	55.5	30.6	13.8
Geography	243	52.7	25.5	21.8
Geology	25	40.0	28.0	32.0
German	130	80.8	17.7	1.5
History	351	53.6	22.2	24.2
Italian	33	81.8	15.2	3.0
Law	102	48.0	14.7	37.3
Mathematics	1031	53.1	25.5	21.4
Media Studies	103	32.0	25.2	42.7
Music	84	84.5	14.3	1.2
Music Technology	34	67.6	23.5	8.8
Philosophy	48	64.6	16.7	18.8
Physical Education	125	58.4	27.2	14.4
Physics	570	61.4	23.0	15.6
Politics	205	54.6	26.3	19.0
Psychology	649	57.8	15.7	26.5

Religious Studies	202	53.5	22.3	24.3
Sociology	443	52.8	19.2	28.0
Spanish	221	66.1	27.1	6.8
Statistics	28	57.1	25.0	17.9

Annex P: AS matched outcomes relative to prediction by subject

Table P.1. AS matched subject outcomes relative to prediction (subjects with more than 500 matched entries)

Subject	Matched students	Matched outcome relative to prediction				
		A	B	C	D	E
Art & Design: Fine Art	582	3.8	3.2	3.9	1.2	0.7
Art & Design: Photography	768	2.3	3.6	4.6	2.9	1.8
Biology	4,452	1.4	2.1	2.8	3.1	2.4
Business Studies	2,439	0.9	1.6	2.2	1.4	1.1
Chemistry	3,843	2.4	3.3	3.8	3.5	2.7
Computing	938	6.0	9.2	10.8	8.5	5.3
Economics	2,011	1.1	2.1	4.3	4.3	2.9
English Language	1,016	0.4	0.3	0.6	0.1	0.1
English Language & Literature	508	1.2	1.3	2.0	0.6	-0.3
English Literature	2,147	1.0	1.7	2.6	0.8	0.2
Further Mathematics	2,990	3.7	4.9	4.7	2.9	1.7
Geography	1,494	0.9	2.3	3.3	2.8	1.4
History	2,525	1.3	2.2	2.1	1.3	0.5
Law	1,243	0.2	0.8	1.6	1.3	0.7
Mathematics	6,172	0.6	1.2	1.3	1.4	1.5
Media Studies	731	0.5	1.0	0.4	0.2	0.4
Physics	2,863	2.6	3.5	4.4	3.8	3.7
Politics	1,069	1.2	2.1	3.1	2.5	1.4
Psychology	5,304	0.9	1.3	1.7	1.8	1.6
Religious Studies	1,243	1.7	1.4	1.8	1.4	1.0
Sociology	3,732	0.3	1.4	2.0	1.8	1.1
Spanish	528	9.6	12.0	12.2	7.8	3.1

Annex Q: Differences between CAGs and final grades by socio-economic group

Annex Q has been published as a separate document.

Annex R: Teacher survey and interviews

To understand the centre assessment grade process, we have collected information from teaching professionals with both a survey and interviews. We wanted to understand the methods used in centres to generate their centre assessment grades, including the sources of evidence used and how their importance varied across centres, and what centres had done to make them as fair as possible. We also wanted to get a sense of individuals' actual experiences, both positive and negative.

We received over 850 full responses and around 70 almost-complete responses to the survey from teaching professionals at all levels of seniority with involvement in at least one part of the centre assessment grade process. Preliminary analysis shows that within GCSE and A level teaching staff, mocks were reported as by far the strongest source of evidence (an average rating of over 80 out of 100) with class tests, class work, and how similar students achieved in previous years also rated over 50 on average. Assignments and the ability of students to perform in exams were next most important source of evidence. The sources of evidence for vocational and technical qualifications included other aspects of student performance such as evidence from completed or partially completed units, together with other coursework or assignments generally rated as more important than mock exams. This is likely to reflect those qualifications which are more coursework focused.

It was reassuring that only around a third of respondents reported any difficulties having effective discussions with colleagues, despite nearly all of the discussions that took place in agreeing grades being remote discussions. Reported difficulties were split between the limitations of remote meeting software and the difficulty of arranging convenient times for discussions.

We asked for ratings of the confidence staff had in their grade judgements on a scale from 0-100 where 100 reflected absolute confidence. The average rating of class teachers for their grades was in the 80s, with nearly all giving ratings over 60. There were a few much lower ratings. When those staff involved in agreeing the final grades for submission to the awarding organisation were asked of their confidence in these grades, which included cases where grades were combined across multiple classes and teachers, the average rated confidence was a little lower.

Finally, we asked all of our survey respondents who had generated grades for individual students to write down 3 words summarising how they felt about the experience. Below is the word cloud we generated from all of the responses:

Figure R.1. Word Cloud of responses to the question ‘Please write down up to three words that summarise how you felt about the experience of generating Centre Assessment Grades’

It is clear that the pressure and stress of this process did affect teaching staff but, conversely, they had confidence in their ability to carry out this task fairly and professionally.

To collect more in-depth qualitative insight, we have interviewed over 50 teaching staff from heads of centre to class teachers in a variety of centre types and qualifications. From these interviews it was clear that although a variety of different processes to generate grades and rank orders had been used, there had been a strong reliance on data to do so. As in the survey responses, the use of mock exam results was important. The teaching staff often reported taking these (or some other test or task taken by all students) as their starting point but refining their final grades and rank orders using their understanding of the effort the students made in the mocks and their effort and progress in the time since the mocks were sat. Mocks (and the other common tests and tasks) were considered a strong evidence to help integrate multiple classes of a subject into a single set of centre assessment grades.

Because many of the centres routinely use data to help predict student grades as part of their continuous improvement, for some the circumstances this year did not present a new challenge. They were used to making predictions and also evaluating the accuracy of their predictions. In these centres there was less need to devise a new centrally-devised approach, more an extension of existing systems to include a final rank order. In other centres teaching staff described how the approach was developed within the senior leadership team and disseminated across all departments to ensure consistency of approach.

One interesting difference between centres was the extent to which prior centre outcomes to guide the profile of centre assessment grades were used. Some centres provided this as a starting point for teachers to aim for, whilst others asked teachers to produce grade and rank order predictions, and then either adjusted them centrally to be better aligned with their historical profiles and/or asked departments to revisit their grades if they were considered too different. Although some centres had departmental meetings in which all teachers contributed and discussed grades and rank orders, a number of class teachers reported that they had no involvement in, or sight of, the final submission – their own class grades and rank order were handed to more senior management who combined classes and might (or might not) have made adjustments to the teacher's grades and rank order.

Almost all teachers we interviewed reported that they expected their predictions to be fair and generally accurate, but they nearly all stated that they had provided centre assessment grades which reflected what student should have been able to achieve in final assessments, but, probably inevitably, not taking into account how every year some students slightly underperform or have a bad day. Together with a desire to err on the optimistic side for students they believed were genuinely right on an important grade borderline, this may explain some of the generosity in the national centre assessment grade profile relative to national outcomes in previous years.

A full analysis of the survey and interviews will follow for publication later in the year.

Published by:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344
public.enquiries@ofqual.gov.uk
www.gov.uk/ofqual

August 2020 **Ofqual/20/6656/1**