CAMBRIDGE
UNIVERSITY PRESS & ASSESSMENT

The impact of the failure to account for KS2 cohort effects on awarding at GCSE

Research Report

Tim Gill

20 October 2025

Author contact details:

Tim Gill
Research Division
Shaftesbury Road
Cambridge
CB2 8EA
UK

tim.gill@cambridge.org
https://www.cambridge.org/

As a department of the university, Cambridge University Press & Assessment is respected and trusted worldwide, managing three world-class examination boards, and maintaining the highest standards in educational assessment and learning. We are a not-for-profit organisation.

Cambridge University Press & Assessment is committed to making our documents accessible in accordance with the WCAG 2.1 Standard. We're always looking to improve the accessibility of our documents. If you find any problems or you think we're not meeting accessibility requirements, contact our team: Research Division
If you need this document in a different format contact us telling us your name, email address and requirements and we will respond within 15 working days.

How to cite this publication:

Gill, T. (2025). *The impact of the failure to account for KS2 cohort effects on awarding at GCSE*. Cambridge University Press & Assessment.

Acknowledgement

This work was undertaken in the Office for National Statistics Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners.

Abstract

What is this report about?

This report investigates a limitation of the current GCSE awarding process in England, which relies on a 'comparable outcomes' approach using Key Stage 2 (KS2) prior attainment data. Specifically, it examines how failing to account for cohort effects – where students perform differently depending on the average ability of their peers – can lead to systematic under- or over-prediction of outcomes for certain awarding organisations (AOs). The research explores whether incorporating centre-level KS2 data into prediction models can improve fairness and accuracy in grade awarding.

What did we do?

We analysed GCSE outcomes in seven subjects (English literature, maths, biology, chemistry, history, religious studies, and music) using data from the National Pupil Database for 2023 and 2024. For each subject, we compared three methods for generating predicted grade distributions:

1. **Prediction matrices** based on current awarding procedures.
2. **Logistic regression excluding centre-level KS2 scores**, using only candidate-level data.
3. **Logistic regression including centre-level KS2 scores**, accounting for both individual and cohort effects.

Candidates from independent/selective schools and those not in Year 11 were excluded, to mirror current awarding practices. We identified AO specifications with candidates who had high or low average KS2 attainment and assessed how predictions differed across methods.

What did we find?

Across all subjects, the logistic regression models *including* centre-level KS2 scores consistently produced more accurate predictions for specifications with high-attaining cohorts. These regression models revealed significant cohort effects: students in higher-attaining centres tended to outperform predictions based solely on individual KS2 scores.

The differences in predicted outcomes were most pronounced at grades 7 and 4, with some specifications seeing shifts of up to 3 percentage points. In contrast, predictions from the logistic regression model *excluding* centre-level data were nearly identical to those from the prediction matrix method, indicating that the observed improvements stemmed from accounting for cohort effects rather than the statistical method itself.

What are the implications?

The findings suggest that the current GCSE awarding process may systematically disadvantage specifications with high-attaining cohorts and advantage those with lower-attaining cohorts. This misalignment could result in grade distributions that are either too harsh or too lenient. Incorporating centre-level KS2 data into prediction models would lead to fairer and more consistent outcomes across specifications. As GCSE awarding in 2026 will again use a "common centres" approach, it is crucial to investigate and correct these discrepancies to ensure equitable standards. We recommend considering the use of logistic regression methods in future awarding processes and reviewing past awards for potential misalignment.

Introduction

The current procedures for awarding GCSE grades in England are based on a 'comparable outcomes' approach. In this approach, the relationship between GCSE grades in a subject and prior attainment (using scores in Key Stage 2 tests) in one year (the 'reference' year) is calculated empirically. In subsequent years, this relationship is used to generate recommended grade boundary marks in that subject ('Statistically Recommended Boundaries', or SRBs). Script judgement is then used to confirm that these outcomes are valid. As such, SRBs are generated using only the relationship between GCSE grades and prior attainment at the candidate level. However, previous research (e.g., Hanushek et al, 2003; Epple & Romano, 2011; Gill, 2018) has shown that performance can also depend on the ability of other students at the school a student attends (so called 'peer' or 'cohort' effects). In particular, the tendency is for students of a given ability to perform better when they are in a school with high attaining students compared with being with low attaining students.

One consequence of using this approach (as outlined in Benton & Sutch, 2014) is that where an awarding organisation (AO) has a much higher proportion of its entries (than other AOs) from candidates in centres with high average prior attainment levels, their outcomes tend to be underpredicted (when compared to predictions based on concurrent attainment[1]). This is likely to be because students in these centres tend to do better at GCSE than the KS2 scores of individuals would predict. This effect is not captured in the current predictions model.

In this research, we investigated an alternative method (using logistic regression models) for making predictions that attempts to account for this underprediction. If this method is successful in doing so, then it could potentially be used in future examination series to adjust outcomes. Moreover, even if AOs prefer not to adopt a logistic regression approach, the current research may identify instances where awards have potentially been too lenient or too harsh so that adjustments can be made.

Data and method

The first step was to identify subjects where there were substantial differences between AO specifications in the ability of the candidates taking the GCSE. We wanted to identify specifications with particularly high (or low) performing candidates in a subject to test the hypothesis that candidates taking these specifications were disadvantaged (or advantaged) by the current awarding process.

The subjects we identified were as follows: English literature, maths, biology, chemistry, history, religious studies (RS), and music. Figures A1 to A7 in Appendix A show the breakdown of candidates into KS2 deciles by specification. These figures all show substantial differences between specifications in the ability of their candidates.

For each of these subjects, we compared the outcomes from three different methods for generating predictions for each specification:

1) Prediction matrices, following as closely as possible the current awarding procedures.

[1] As measured by mean GCSE

2) Logistic regression excluding centre level mean KS2. i.e. only accounting for candidate level KS2 scores. This allows us to see the impact of a change of statistical methodology without accounting for centre level effects.
3) Logistic regression including centre level mean KS2. i.e. accounting for candidate level KS2 scores and centre level mean KS2 scores. This allows us to see the impact of the methodology and also accounting for centre level effects.

The use of the logistic regression method (including centre level mean KS2) was found in previous research to be successful in reducing the under-prediction issue (Benton & Sutch, 2014). The prediction methods are outlined below.

Prior to undertaking the analysis, we excluded some candidates in order to mirror the current procedures for awarding grades. Specifically, we excluded candidates who were younger than 16, those attending independent or selective schools, and those with no GCSE results.

Prediction matrix method

For this method, the prediction matrix was generated using 2023 data and was then applied to the 2024 data. This follows the current awarding procedure of generating a prediction matrix in a reference year and the applying it to the current year.

The first part of the process for the prediction matrix method was to allocate candidates to KS2 deciles for 2023 and 2024, based on their performance in the KS2 tests. This was done by normalising the scaled scores for both the KS2 Maths and Reading tests. These new scores were then combined and normalised again. For each of the normalisations, the percentile rank of each candidate was calculated and converted to the equivalent point on a normal distribution with a mean of 50 and standard deviation of 16.67. The normalised score was then used to split candidates into deciles. This process was done separately for candidates in the 2023 data and the 2024 data. Note that deciles were defined once in each year, based on all GCSE candidates included in analysis, rather than separately for each subject.

For each subject investigated, a prediction matrix was then generated using the 2023 data. A prediction matrix is a table which shows the relationship between KS2 deciles and GCSE grades. It shows, for each KS2 decile, the cumulative percentage of candidates achieving each GCSE grade (or higher). Table 1 gives a hypothetical example of a prediction matrix. This shows that, for example, there were 30,240 candidates in the highest KS2 decile taking this GCSE subject. Of these, 26.38% achieved a GCSE grade 9, 52.01% a GCSE grade 8 (or higher) and so on.

For generating the prediction matrices, we excluded candidates in centres which had fewer than 5 candidates taking the subject in question. This was to allow straightforward comparisons with the logistic regression approach (described below).

The next step was to ‘apply’ this prediction matrix to the 2024 data. This was done for each specification separately for the cohort of candidates taking the qualification and with known KS2 scores. The number of candidates in each KS2 decile in 2024 was calculated and then, using the percentages in the prediction matrix, the number in each decile expected to achieve each GCSE grade was generated. Table 2 gives an example of the calculation for KS2 decile 1 for one specification, where there were 16,156 candidates in the 2024 data. The expected number of decile 1 candidates achieving a grade 9 is 26.38% of 16,156 which is 4,262.

The sum of these numbers across all KS2 deciles gives the overall numbers expected to achieve each grade for that specification.

Table 1: Hypothetical example of a prediction matrix for a GCSE subject

KS2 decile	No of candidates	% of candidates	9	8	7	6	5	4	3	2	1	0
1 (highest)	30,240	22.7	26.38	52.01	75.67	88.88	97.35	99.25	99.65	99.88	99.95	100.00
2	26,122	19.6	10.05	28.08	50.99	76.26	93.50	98.75	99.55	99.77	99.88	100.00
3	21,008	15.8	6.98	20.25	38.88	66.45	87.26	96.75	99.1	99.54	99.65	100.00
4	16,267	12.2	4.22	13.67	25.98	55.19	80.73	94.00	98.5	99.26	99.44	100.00
5	12,678	9.5	1.85	8.27	20.26	43.81	71.24	89.92	97.47	99.05	99.33	100.00
6	9,555	7.2	1.08	4.44	12.68	31.29	59.07	80.96	93.38	97.79	99.02	100.00
7	6,805	5.1	0.98	3.12	8.91	23.87	48.12	73.28	91.85	96.81	98.54	100.00
8	4,999	3.8	0.37	1.85	5.54	14.87	35.87	60.35	85.55	95.67	98.54	100.00
9	3,502	2.6	0.14	0.84	2.51	7.67	20.64	41.45	73.51	92.49	98.41	100.00
10 (lowest)	1,975	1.5	0.08	0.34	0.82	1.87	4.99	11.87	39.25	71.95	97.24	100.00
Total	133,151	100	9.90	23.55	40.08	60.80	79.53	90.31	96.23	98.58	99.50	100.00

Table 2: Hypothetical calculation of cumulative number of candidates in KS2 decile 1 expected to achieve each grade (for one specification)

KS2 decile	No of candidates	9	8	7	6	5	4	3	2	1	0
1	16,156	4,262	8,403	12,225	14,360	15,728	16,035	16,100	16,137	16,148	16,156

Logistic regression excluding centre mean KS2 scores method

For the logistic regression analysis, we fitted models on the data from 2023 predicting the probability of a candidate achieving a particular grade or higher in a subject (for key grades 1, 4, 7, and 9). In each model, the only predictor variable was the candidate level KS2 normalised score (as explained above). The regressions were fitted using data from all specifications together. As with the prediction matrix method, candidates in centres with fewer than 5 entries in the subject were excluded.

The parameter estimates for the variables included in the models were then used to generate predicted probabilities of achieving each key grade (or higher) for all candidates in 2024. The overall expected percentage achieving each key grade (or higher) was then just the average probability for that grade across all candidates (within the specification). This generated expected percentages for each specification separately.

Logistic regression including centre mean KS2 scores method

As above, but with an additional predictor variable of the centre level mean KS2 score. This was calculated as the mean of the KS2 normalised score in each centre. As with the other logistic regression, these regressions were fitted using data from all specifications together. Again, candidates in centres with fewer than 5 entries in the subject were excluded.

The outcomes from these alternative prediction methods were compared with the actual outcomes in the data (in 2024).

NPD data

The data for the analyses came from the Key Stage 4 (KS4) extracts of the National Pupil Database (NPD) for 2023 and 2024. The NPD is administered by the Department for Education (DfE) and includes examination results for all students in all qualifications and subjects in schools and colleges in England, as well as student and school background characteristics such as gender, ethnicity, level of income-related deprivation and school type.

The NPD data includes the following variables which were required for the analysis:

- Qualification and subject – to identify the GCSE subjects we were interested in.
- Final grade achieved by candidate – required to run the logistic regression analysis, to create the prediction matrices, and to compare these against the 2024 actual outcomes.
- Specification reference number – to identify each specification.
- KS2 scores for each candidate – in Reading and Maths.
- Centre reference number – to group candidates into centres so as to calculate the centre-level mean KS2 scores.
- School type – to mimic the current approach to awarding grades, we excluded candidates attending independent or selective schools.
- Year group – again, to mimic the current approach to awarding grades, we excluded candidates not in year 11.

In all the analyses, specifications were anonymised so that AOs cannot be identified. The specifications included came from the four main AOs for GCSEs (AQA, Edexcel, OCR, and Eduqas). For some subjects there were more than four specifications, this was because some AOs offered more than one specification in the subject.

Results

English literature

Figure A1 in Appendix A shows the breakdown of the students taking English literature into KS2 deciles by specification. This shows that spec 3 had students with much higher ability levels than other specifications. This means that we expected predictions to be relatively higher using the logistic regression model (with centre mean KS2 score as a predictor) for spec 3 compared with other specifications.

Tables B1 and B2 in Appendix B show the results of the two sets of regression models, one without and one with centre mean KS2 score as a predictor. Separate models were run for each of the key GCSE grades (1, 4, 7, 9). Table B2 shows that, the parameter estimate for the centre mean KS2 score was positive and highly significant ($p<0.001$) in each model. This suggests that being in a higher attaining centre was associated with better performance, and also that including this variable improved the model.

Table 3 and Figures 1 to 3 compare the predictions of the cumulative percentages achieving each key grade from the different methods. For grade 1, the percentages were almost identical from each method, so they are not shown graphically[2].

Table 3: Predictions of cumulative percentages achieving key grades from different methods (English literature)

Specification	Method	9	7	4	1
Spec 1	Pred mat	3.0	18.6	73.7	98.2
	Log. reg. 1	3.0	18.6	73.6	98.2
	Log. reg. 2	2.9	18.4	73.5	98.2
	Actual	2.9	18.1	73.3	98.0
Spec 2	Pred mat	3.0	18.7	73.8	98.2
	Log. reg. 1	3.0	18.7	73.8	98.3
	Log. reg. 2	3.1	18.8	73.6	98.2
	Actual	2.8	17.8	74.4	98.2
Spec 3	Pred mat	3.6	21.4	77.0	98.5
	Log. reg. 1	3.8	21.7	77.1	98.5
	Log. reg. 2	4.7	23.9	78.9	98.8
	Actual	3.3	20.1	76.1	96.6
Spec 4	Pred mat	2.9	18.3	73.3	98.2
	Log. reg. 1	2.9	18.2	73.3	98.2
	Log. reg. 2	2.8	17.7	72.9	98.2
	Actual	3.0	18.7	72.8	97.6

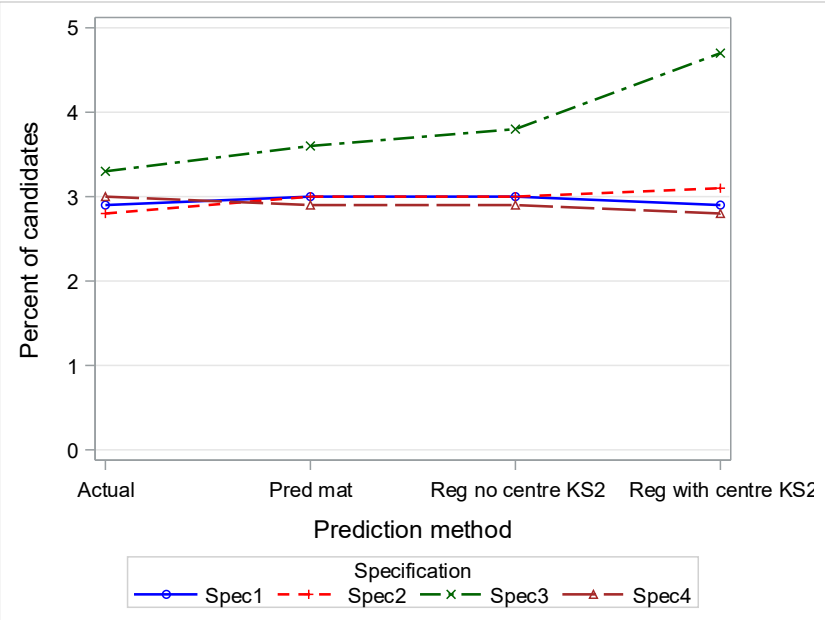

Figure 1: Comparison of predicted cumulative percentages, by method and specification (English literature, grade 9)

[2] This was the case for all other subjects also

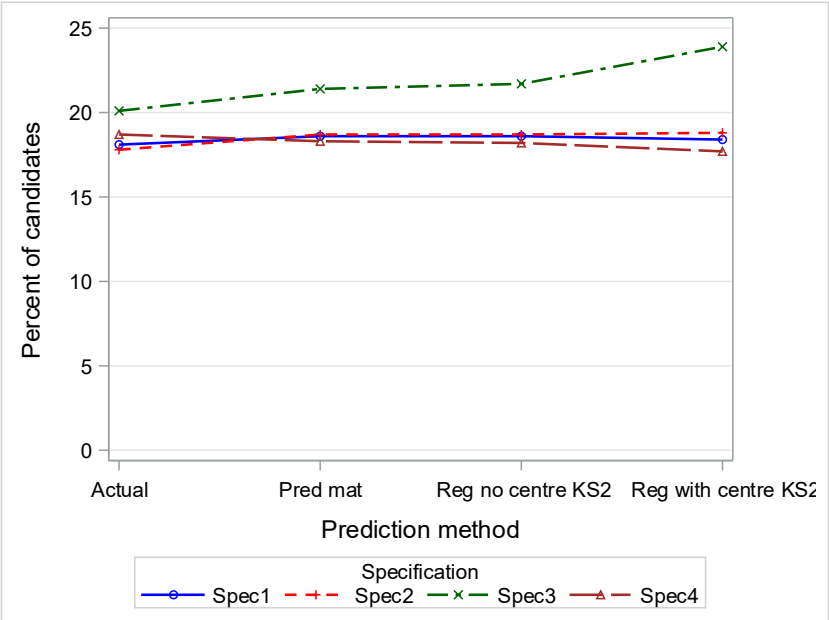

Figure 2: Comparison of predicted cumulative percentages, by method and specification (Englis literature, grade 7)

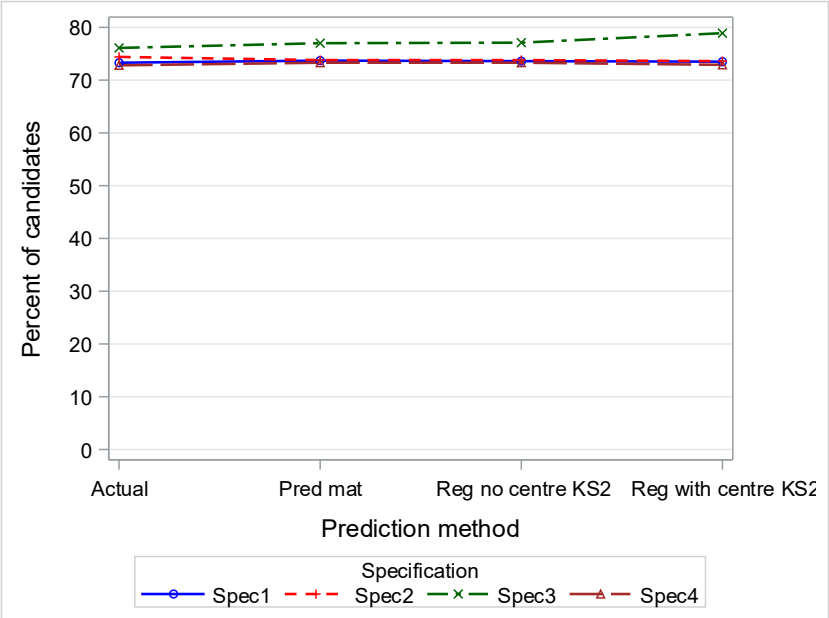

Figure 3: Comparison of predicted cumulative percentages, by method and specification (Eng. Lit., grade 4)

What we were most interested in here was the difference in percentages between the current method (prediction matrix) and the logistic regression with centre mean KS2 score. At each grade, this difference was much larger for spec 3 than for the other specifications. For spec 3 the differences were 1.1 percentage points at grade 9, 2.5 percentage points at grade 7, and 1.9 percentage points at grade 4. For the other specifications the differences were very small, and the percentages were all higher for the prediction matrix method.

It is also worth noting that the differences between the prediction matrix method and the logistic regression without centre mean KS2 method were very small at each grade, which

suggests the difference between prediction matrix method and logistic regression with centre mean KS2 was due to the inclusion of centre mean KS2, not the method (logistic regression) itself.

Maths

Figure A2 in Appendix A shows the percentage of candidates in each KS2 decile by specification for maths. This shows that spec 3 and spec 4 had lower attaining students than other specifications.

Tables B3 and B4 in Appendix B show the results of the two sets of regression models, Table B4 shows that, the centre mean KS2 score was highly significant ($p<0.001$) in each model, which suggests that including this variable improved the model.

Table 4 and Figures 4 to 6 compare the predictions of the cumulative percentages achieving each key grade from the different methods.

This shows that there was almost no difference between prediction matrix and logistic regression outcomes for spec 1, spec 2, and spec 3. However, for spec 4, at grades 7 and 4 the percentages were lower for the logistic regression with centre mean KS2 method than for prediction matrix method, by 1.2 pp and 1.9 pp respectively. This was likely to be because of spec 4 having lower ability candidates than other specifications.

Table 4: Predictions of cumulative percentages achieving key grades from different methods (maths)

Specification	Method	9	7	4	1
Spec 1	Pred mat	3.4	19.6	73.1	98.6
	Log. reg. 1	3.3	19.5	73.1	98.6
	Log. reg. 2	3.3	19.6	73.3	98.6
	Actual	3.4	19.6	73.3	98.7
Spec 2	Pred mat	3.3	19.3	72.9	98.6
	Log. reg. 1	3.3	19.2	72.8	98.6
	Log. reg. 2	3.3	19.2	72.8	98.6
	Actual	3.2	19.1	72.3	98.5
Spec 3	Pred mat	2.8	17.0	70.0	98.4
	Log. reg. 1	2.8	17.0	69.8	98.4
	Log. reg. 2	2.7	16.8	69.7	98.4
	Actual	2.9	17.7	70.6	98.6
Spec 4	Pred mat	2.2	14.7	67.0	98.3
	Log. reg. 1	2.3	14.7	67.0	98.2
	Log. reg. 2	2.1	13.5	65.1	97.9
	Actual	2.1	11.4	60.3	96.0

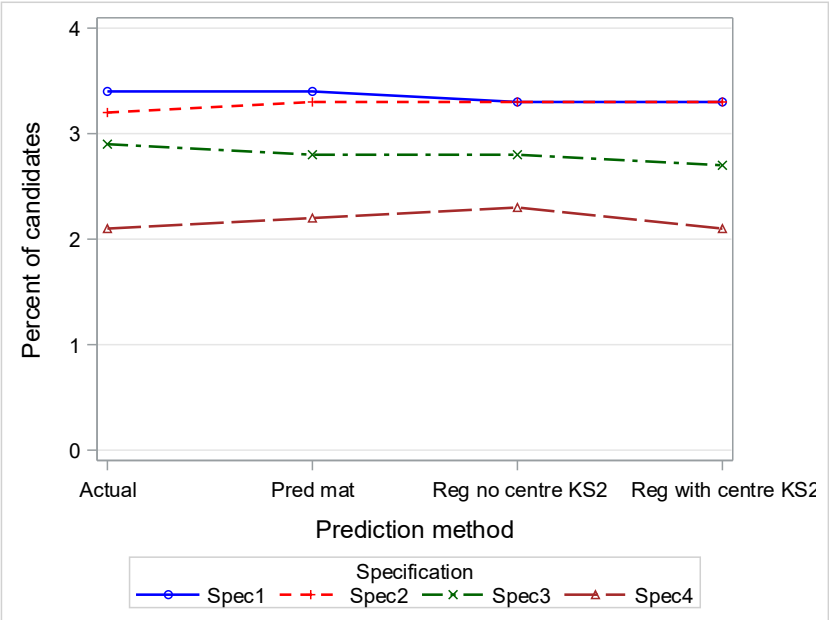

Figure 4: Comparison of predicted cumulative percentages, by method and specification (maths, grade 9)

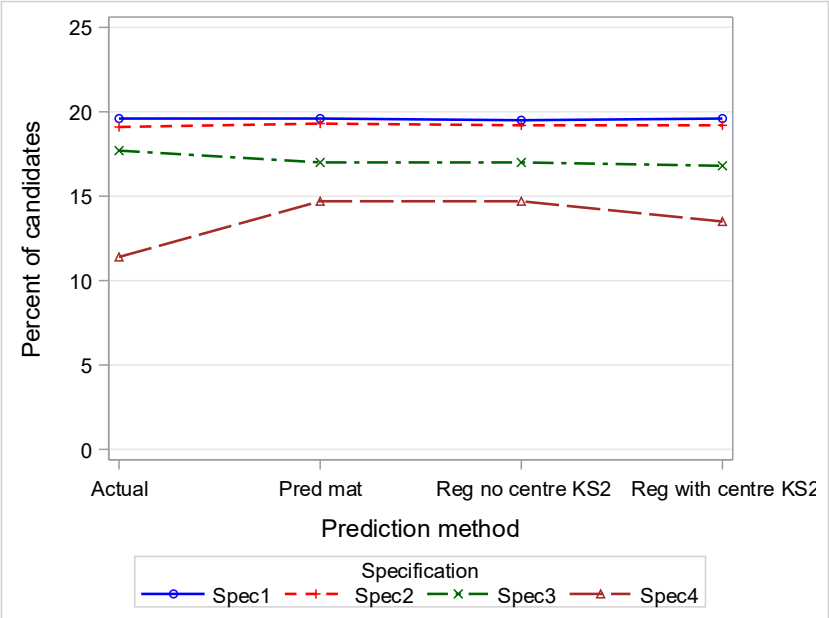

Figure 5: Comparison of predicted cumulative percentages, by method and specification (maths, grade 7)

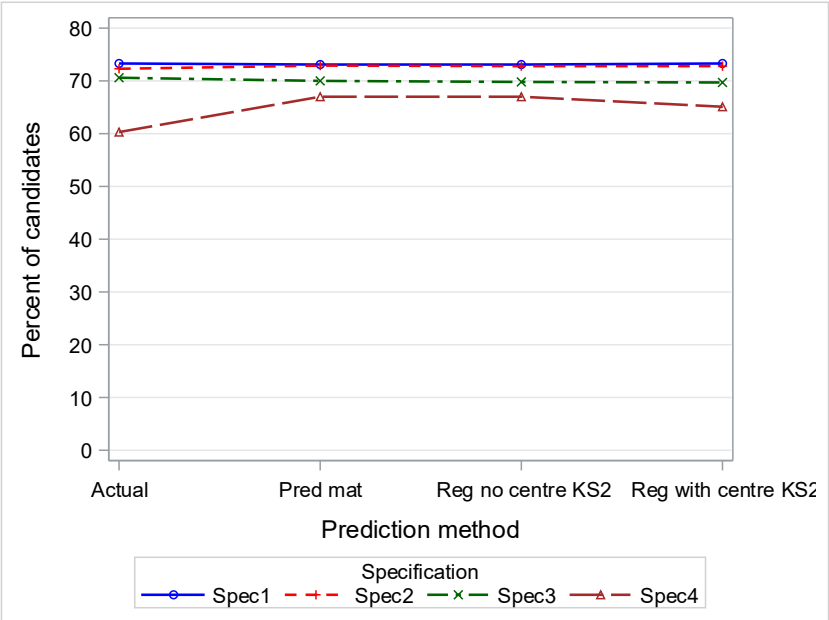

Figure 6: Comparison of predicted cumulative percentages, by method and specification (maths, grade 4)

Biology

Figure A3 in Appendix A shows the percentage of candidates in each KS2 decile by specification for biology. This shows that spec 3 had the highest attaining students, slightly higher than spec 2. Spec 4 had the lowest attaining students.

Tables B5 and B6 in Appendix B show the results of the two sets of regression models, Table B6 shows that, the centre mean KS2 score was highly significant ($p<0.001$) in each model.

Table 5 and Figures 7 to 9 compare the predictions of the cumulative percentages achieving each key grade from the different methods. This shows that only the spec 3 percentages were higher for the regression with centre mean KS2 score method than for the prediction matrix method. At grade 9, this difference was 0.9 pp and at grade 7 it was 1.7 pp. There was almost no difference at grade 4 for spec 3.

Although spec 4 had the lowest attaining students, there was no evidence of lower percentages using the logistic regression method than using the prediction matrix method.

Table 5: Predictions of cumulative percentages achieving key grades from different methods (biology)

Specification	Method	9	7	4	1
Spec 1	Pred mat	9.6	37.6	89.0	99.4
	Log. reg. 1	9.6	37.5	89.0	99.4
	Log. reg. 2	9.5	37.5	89.0	99.4
	Actual	9.7	37.5	89.2	99.6
Spec 2	Pred mat	10.0	38.8	90.2	99.5
	Log. reg. 1	10.0	38.8	90.1	99.5
	Log. reg. 2	10.0	38.8	90.2	99.5
	Actual	10.2	39.1	90.0	99.2
Spec 3	Pred mat	10.3	39.5	90.5	99.4
	Log. reg. 1	10.5	39.6	90.5	99.5
	Log. reg. 2	11.2	41.2	90.7	99.5
	Actual	9.9	37.9	90.8	99.5
Spec 4	Pred mat	8.0	32.5	84.9	99.2
	Log. reg. 1	7.9	32.5	84.8	99.2
	Log. reg. 2	7.8	32.5	85.5	99.3
	Actual	8.8	32.9	84.5	99.4

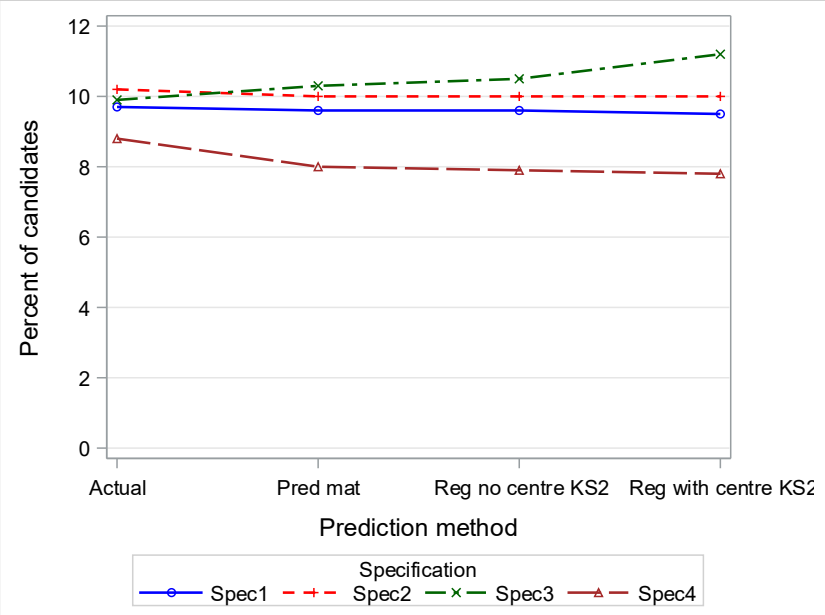

Figure 7: Comparison of predicted cumulative percentages, by method and specification (biology, grade 9)

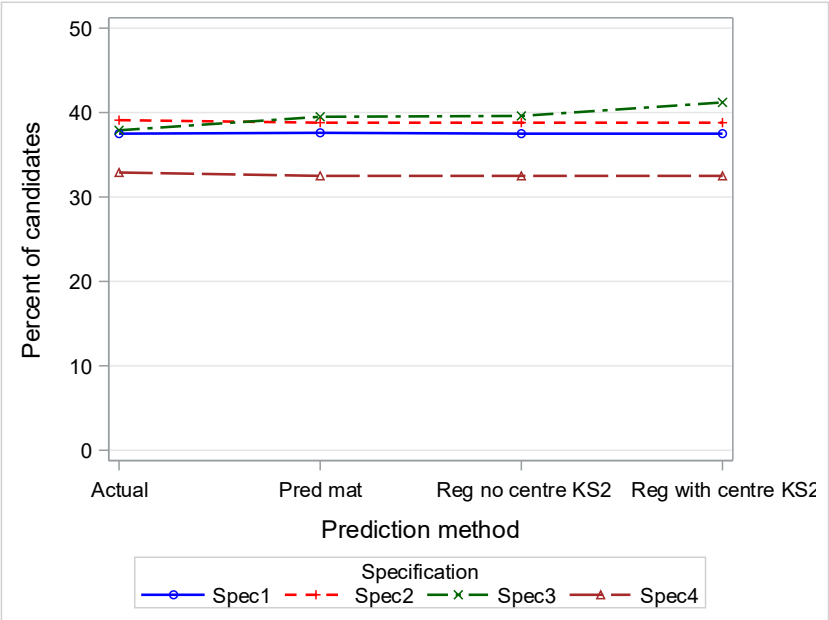

Figure 8: Comparison of predicted cumulative percentages, by method and specification (biology, grade 7)

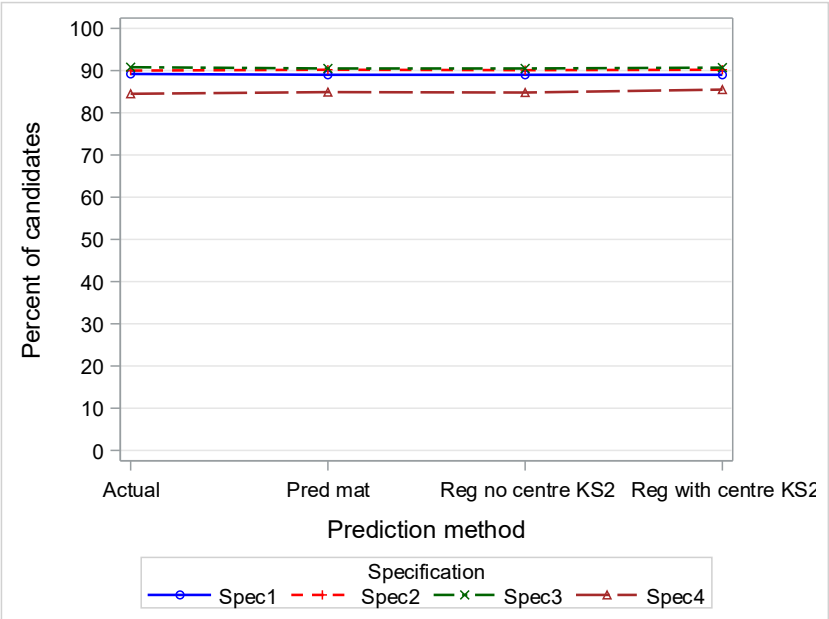

Figure 9: Comparison of predicted cumulative percentages, by method and specification (biology, grade 4)

Chemistry

Figure A4 in Appendix A shows the percentage of candidates in each KS2 decile by specification for chemistry. The pattern was very similar to biology, with spec 3 having the highest attaining students, slightly higher than spec 2 and spec 1. Spec 4 had the lowest attaining students.

Tables B7 and B8 in Appendix B show the results of the two sets of regression models, Table B8 shows that, the centre mean KS2 score was highly significant ($p<0.001$) in each model.

Table 6 and Figures 10 to 12 compare the predictions of the cumulative percentages achieving each key grade from the different methods. The results were very similar to biology, with higher percentages for spec 3 for the regression with centre mean KS2 score method than for the prediction matrix method. At grade 9, this difference was 0.8 pp and at grade 7 it was 1.8 pp.

Table 6: Predictions of cumulative percentages achieving key grades from different methods (chemistry)

Specification	Method	9	7	4	1
Spec 1	Pred mat	9.8	38.4	88.3	99.2
	Log. reg. 1	9.8	38.3	88.2	99.2
	Log. reg. 2	9.8	38.2	88.3	99.2
	Actual	10.2	39.0	89.2	99.6
Spec 2	Pred mat	10.1	39.0	88.8	99.2
	Log. reg. 1	10.1	39.0	88.7	99.3
	Log. reg. 2	10.1	39.0	88.8	99.3
	Actual	9.7	39.0	88.9	99.2
Spec 3	Pred mat	10.6	40.2	89.8	99.3
	Log. reg. 1	10.7	40.3	89.9	99.3
	Log. reg. 2	11.4	42.0	90.2	99.3
	Actual	9.6	39.5	89.4	99.3
Spec 4	Pred mat	8.0	33.1	83.7	99.0
	Log. reg. 1	8.0	33.1	83.6	99.0
	Log. reg. 2	7.9	33.0	84.0	99.1
	Actual	8.5	33.1	83.1	99.1

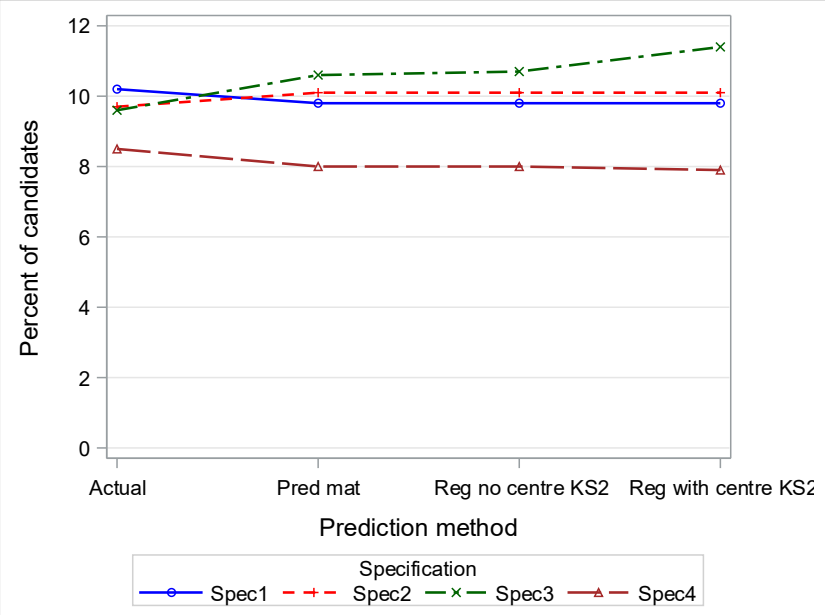

Figure 10: Comparison of predicted cumulative percentages, by method and specification (chemistry, grade 9)

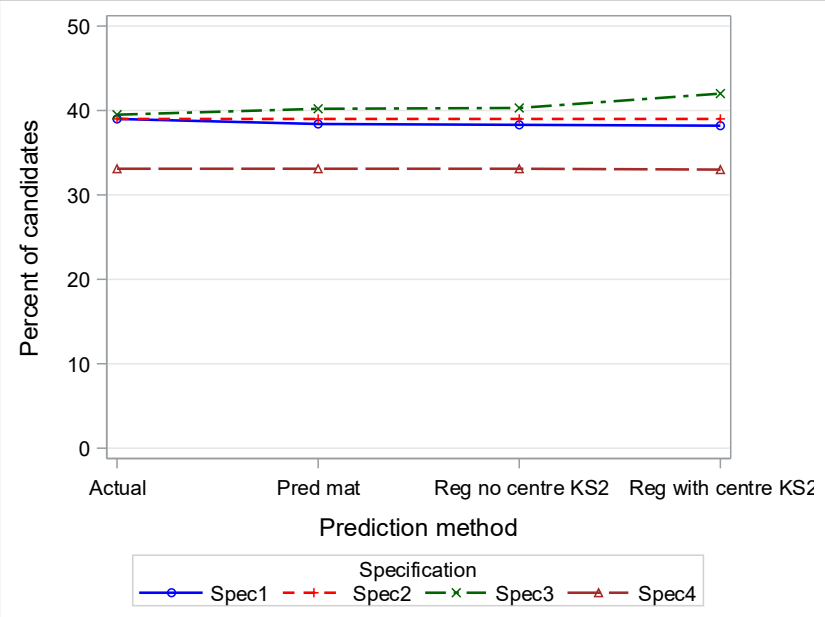

Figure 11: Comparison of predicted cumulative percentages, by method and specification (chemistry, grade 7)

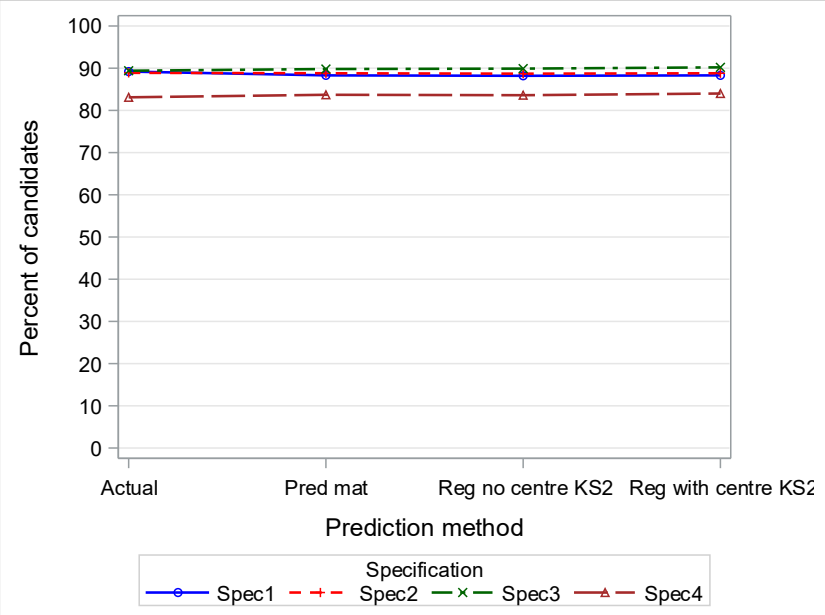

Figure 12: Comparison of predicted cumulative percentages, by method and specification (chemistry, grade 4)

History

Figure A5 in Appendix A shows that spec 3 had substantially higher ability students than any other specification. It also shows that spec 5 had slightly lower ability students than other specifications.

Table B10 in Appendix B shows that the parameter estimate for the centre mean KS2 score was positive and highly significant ($p<0.001$) in each model.

Table 7 and Figures 13 to 15 compare the predictions of the cumulative percentages achieving each key grade from the different methods. At each grade, the percentage for spec 3 was higher for the logistic regression with centre mean KS2 than for the prediction matrix. The differences were 0.9 pp at grade 9, 3.0 pp at grade 7, 2.8 pp at grade 4, and 0.5 pp at grade 1. In contrast, for spec 5, the percentages were lower for the logistic regression than for the prediction matrix (by 0.2 pp at grade 9, 0.8 pp at grade 7, 1.3 pp at grade 4, and 0.3 pp at grade 1).

Table 7: Predictions of cumulative percentages achieving key grades from different methods (history)

Specification	Method	9	7	4	1
Spec 1	Pred mat	4.7	23.6	62.6	96.7
	Log. reg. 1	4.6	23.6	62.6	96.7
	Log. reg. 2	4.7	23.9	63.0	96.8
	Actual	4.7	23.8	62.9	97.0
Spec 2	Pred mat	4.4	22.8	61.6	96.6
	Log. reg. 1	4.4	22.7	61.6	96.6
	Log. reg. 2	4.3	22.6	61.4	96.5
	Actual	4.8	23.2	62.0	96.1
Spec 3	Pred mat	5.8	27.0	65.8	97.0
	Log. reg. 1	5.8	27.2	66.0	97.0
	Log. reg. 2	6.7	30.0	68.6	97.5
	Actual	5.1	26.4	65.7	97.4
Spec 4	Pred mat	4.3	22.4	61.1	96.5
	Log. reg. 1	4.3	22.4	61.1	96.5
	Log. reg. 2	4.2	22.1	60.7	96.4
	Actual	4.6	23.2	61.6	96.4
Spec 5	Pred mat	4.1	21.5	60.1	96.4
	Log. reg. 1	4.1	21.5	60.0	96.3
	Log. reg. 2	3.9	20.7	58.8	96.1
	Actual	3.9	21.6	60.4	94.1

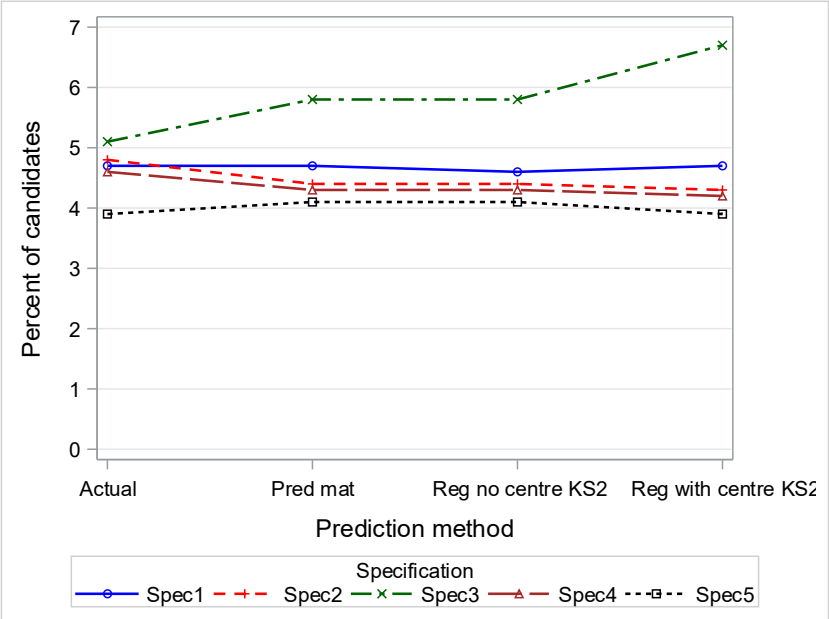

Figure 13: Comparison of predicted cumulative percentages, by method and specification (history, grade 9)

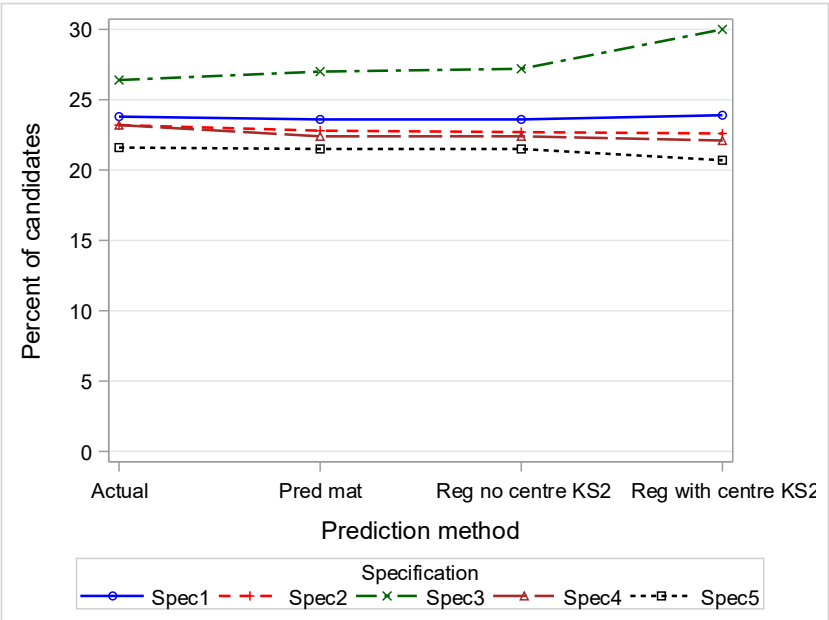

Figure 14: Comparison of predicted cumulative percentages, by method and specification (history, grade 7)

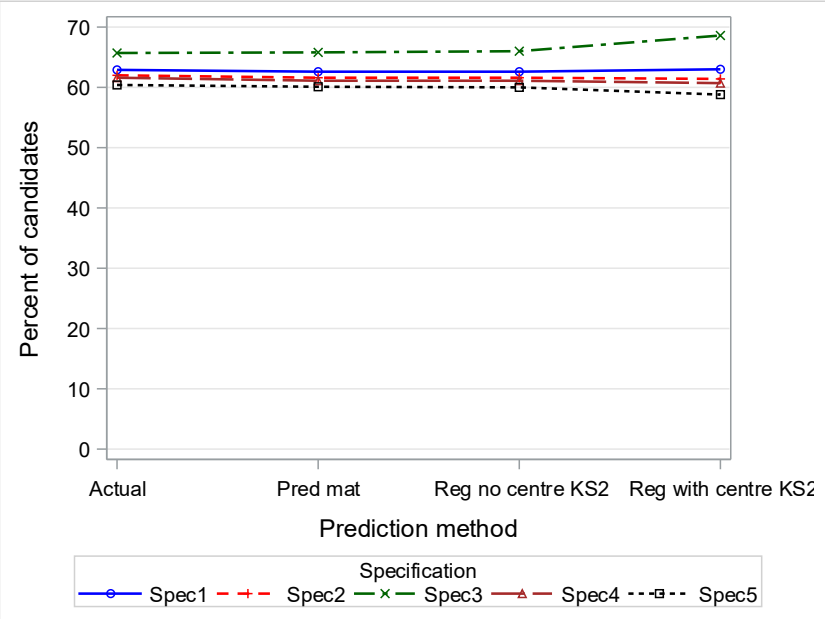

Figure 15: Comparison of predicted cumulative percentages, by method and specification (history, grade 4)

Religious Studies

Figure A6 in Appendix A shows that spec 2 and spec 5 had higher ability students than the other specifications. Table B12 in Appendix B shows that the centre mean KS2 score variable was positive and highly significant in all the models.

Table 8 and Figures 16 to 18 compare the predictions of the cumulative percentages achieving each key grade from the different methods. This shows that both spec 5 and spec 2 had similar increases in percentages using the logistic regression with centre mean KS2 score method compared with using the prediction matrix method. The differences for spec 5 and spec 2 were, respectively, 0.9 pp and 0.6 pp at grade 9, 1.7 pp and 1.9 pp at grade 7, and 1.2 pp and 1.9 pp at grade 4.

Table 8: Predictions of cumulative percentages achieving key grades from different methods (RS)

Specification	Method	9	7	4	1
Spec 1	Pred mat	6.0	26.9	70.3	98.5
	Log. reg. 1	6.0	26.8	70.3	98.5
	Log. reg. 2	6.0	26.8	70.2	98.5
	Actual	5.8	26.8	70.0	97.9
Spec 2	Pred mat	6.9	29.6	73.4	98.7
	Log. reg. 1	6.9	29.6	73.4	98.7
	Log. reg. 2	7.5	31.5	75.3	98.9
	Actual	7.1	30.2	73.9	97.9
Spec 3	Pred mat	5.9	26.6	70.1	98.5
	Log. reg. 1	5.9	26.5	70.1	98.4
	Log. reg. 2	5.8	26.5	70.1	98.5
	Actual	6.4	28.2	70.3	98.2
Spec 4	Pred mat	6.3	27.5	70.7	98.5
	Log. reg. 1	6.3	27.5	70.7	98.5
	Log. reg. 2	6.4	27.8	70.9	98.5
	Actual	6.6	27.9	70.7	98.4
Spec 5	Pred mat	6.8	29.1	72.2	98.6
	Log. reg. 1	7.0	29.3	72.4	98.6
	Log. reg. 2	7.7	30.8	73.4	98.7
	Actual	6.5	28.6	71.8	98.7
Spec 6	Pred mat	5.8	26.3	69.8	98.4
	Log. reg. 1	5.8	26.2	69.8	98.4
	Log. reg. 2	5.6	25.9	69.5	98.4
	Actual	5.7	26.2	69.7	98.4

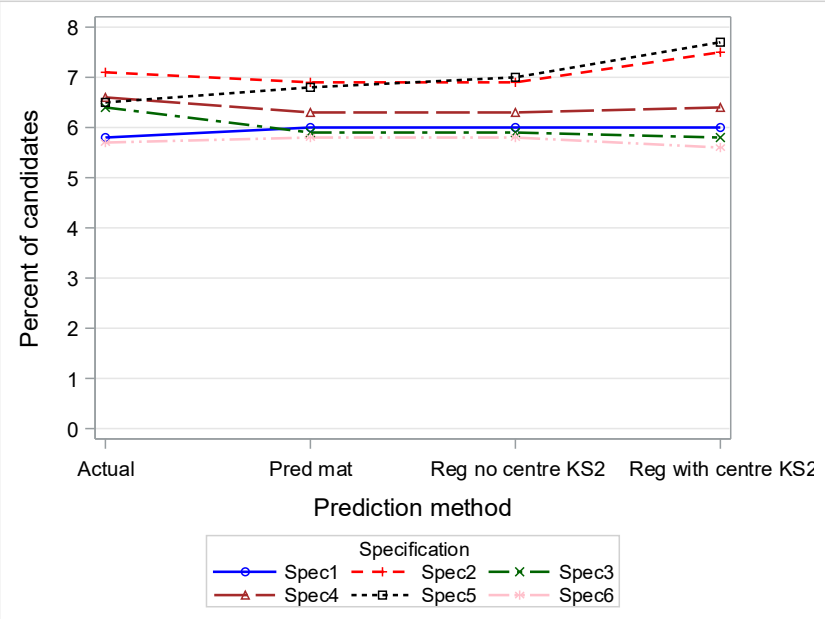

Figure 16: Comparison of predicted cumulative percentages, by method and specification (RS, grade 9)

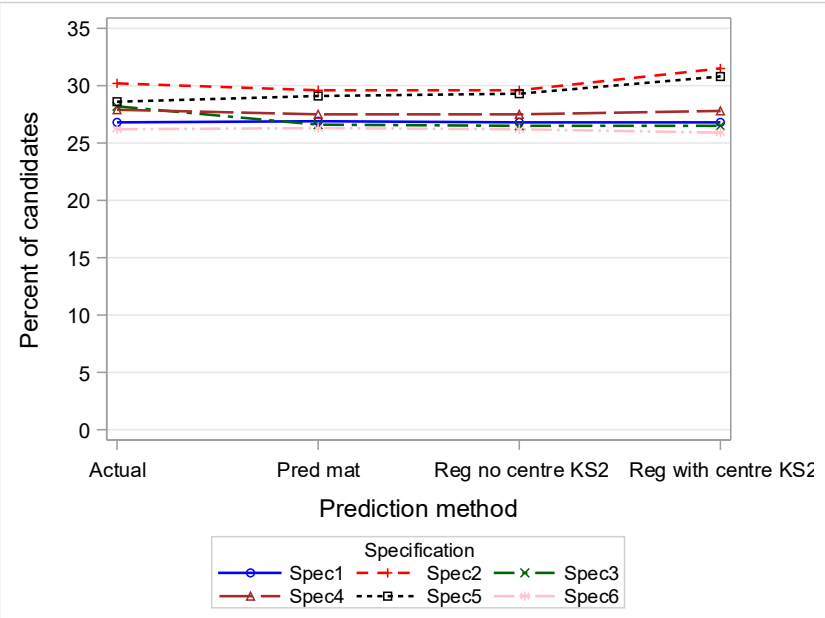

Figure 17: Comparison of predicted cumulative percentages, by method and specification (RS, grade 7)

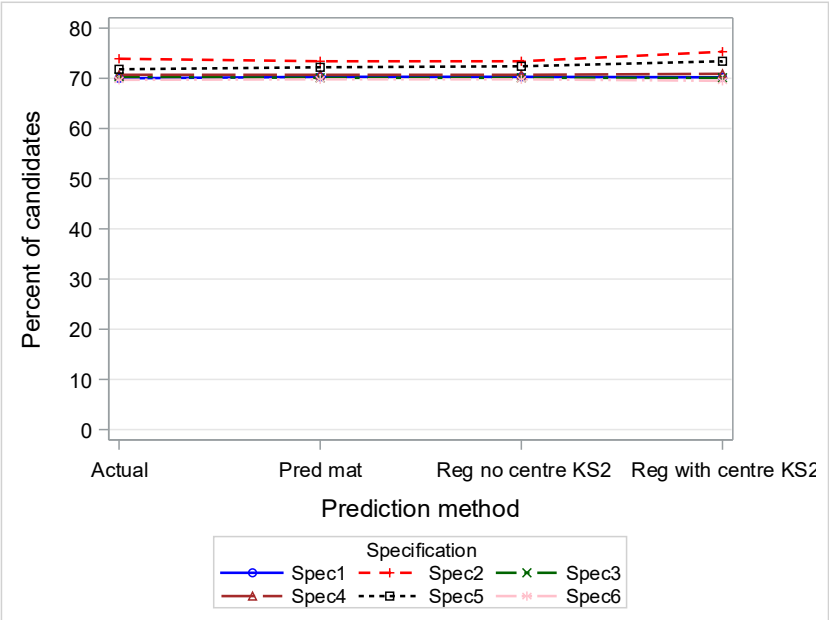

Figure 18: Comparison of predicted cumulative percentages, by method and specification (RS, grade 4)

Music

Figure A7 in Appendix A shows that spec 2 had students with substantially higher ability than other specifications. In contrast, spec 3 and spec 4 had lower ability students. Table B14 shows that the centre mean KS2 score was highly significant in the logistic regression models.

Table 9 and Figures 19 to 21 compare the predictions of the cumulative percentages achieving each key grade from the different methods. At each grade, using the logistic regression with centre mean KS2 method led to higher percentages than using the prediction matrix method for spec 2. This amounted to 0.9 pp at grade 9, 2.4 pp at grade 7, and 1.7 pp at grade 4.

The reverse was true for spec 1 and spec 3, with lower percentages from using the logistic regression with centre mean KS2 method than using the prediction matrix method (0.8 pp and 0.5 pp at grade 9, 2.4 pp and 1.5 pp at grade 7, and 2.1 pp and 1.3 pp at grade 4)

Table 9: Predictions of cumulative percentages achieving key grades from different methods (music)

Specification	Method	9	7	4	1
Spec 1	Pred mat	5.4	25.4	71.3	98.1
	Log. reg. 1	4.9	23.9	69.9	98.0
	Log. reg. 2	4.6	23.0	69.1	97.9
	Actual	5.2	25.3	71.7	98.5
Spec 2	Pred mat	6.5	28.9	75.4	98.6
	Log. reg. 1	6.5	29.0	75.4	98.5
	Log. reg. 2	7.4	31.3	77.1	98.6
	Actual	6.4	28.4	76.1	98.3
Spec 3	Pred mat	4.7	23.2	69.0	97.9
	Log. reg. 1	4.7	23.0	68.9	97.9
	Log. reg. 2	4.2	21.7	67.7	97.7
	Actual	5.1	23.3	70.1	97.8
Spec 4	Pred mat	5.0	24.1	70.0	98.0
	Log. reg. 1	5.4	25.3	71.3	98.1
	Log. reg. 2	5.2	24.9	70.9	98.1
	Actual	5.0	23.9	70.0	98.1

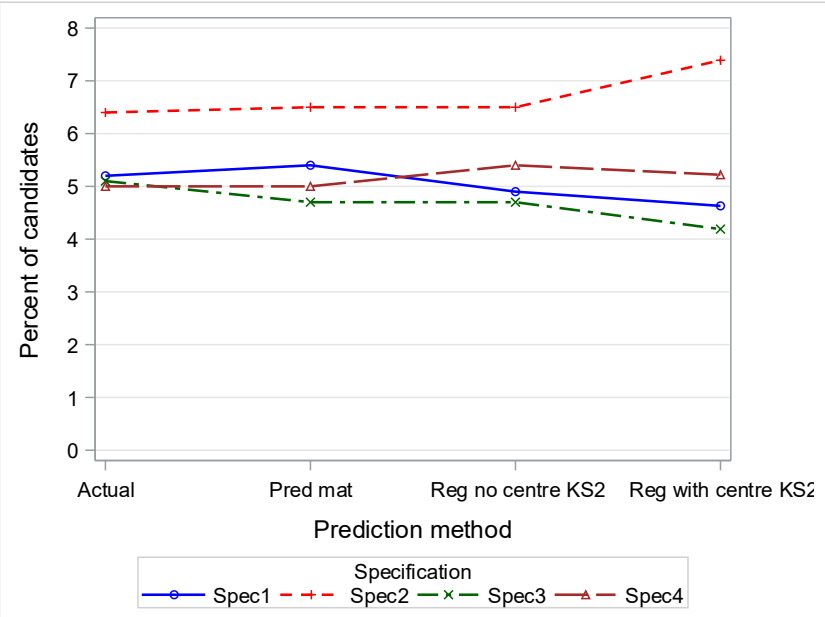

Figure 19: Comparison of predicted cumulative percentages, by method and specification (music, grade 9)

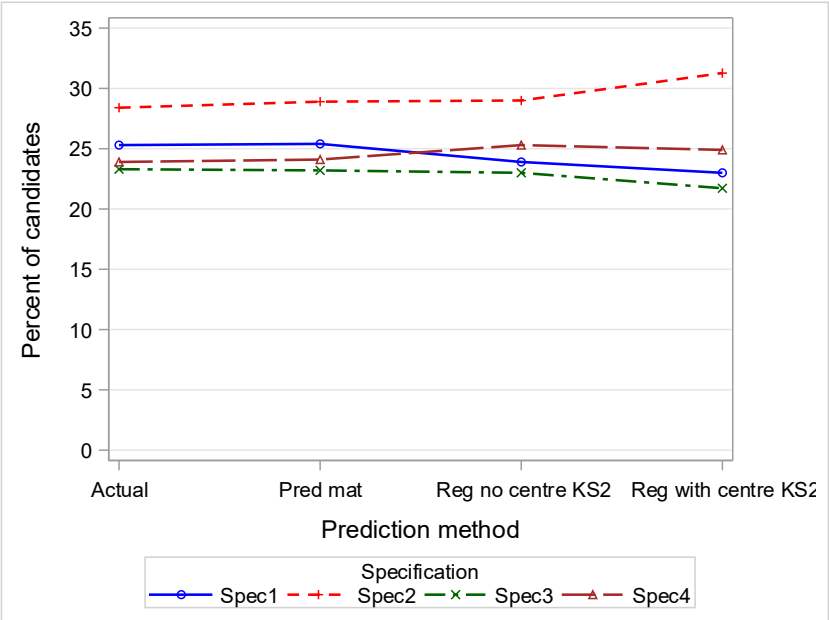

Figure 20: Comparison of predicted cumulative percentages, by method and specification (music, grade 7)

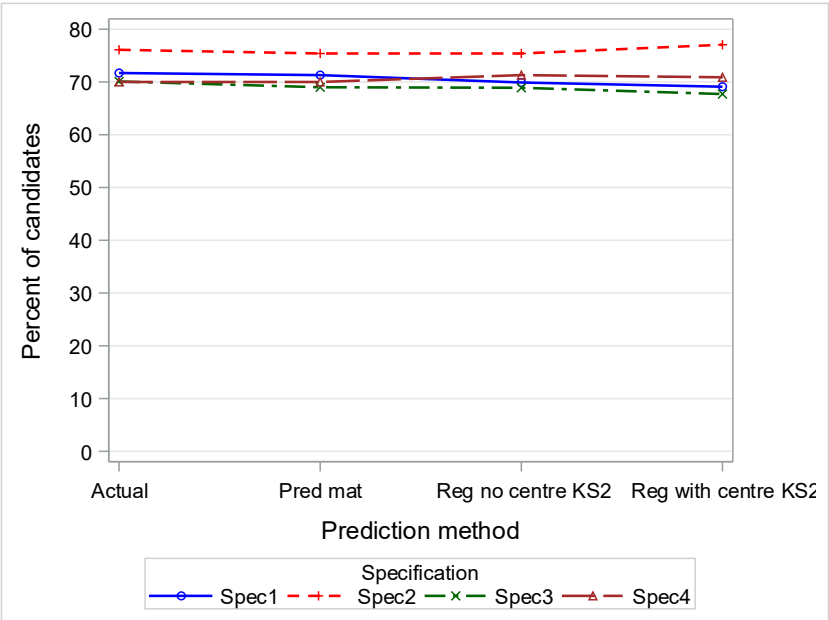

Figure 21: Comparison of predicted cumulative percentages, by method and specification (music, grade 4)

Conclusion

In the analysis presented here, we compared the grade predictions from the current GCSE awarding process with predictions from a possible alternative method using a logistic regression which accounted for the fact that students tend to perform better in higher attaining centres.

We identified a number of subjects where at least one AO specification had higher proportions of high attaining candidates than other specifications and therefore might be underpredicted (compared with predictions from concurrent attainment) using the current awarding process.

The logistic regression models consistently identified significant cohort effects where, even with equal prior attainment, students at schools where average prior attainment was high tended to perform better. As such, using logistic regression instead of the current awarding method led to increases in the predicted percentages achieving key grades (or higher) for specifications with high proportions of candidates in high attaining centres. This was the case across most grades and across all subjects we investigated. The differences were small but not insignificant, particularly at grades 7 and 4. The largest difference was 3 percentage points at grade 7 for one specification in history. Furthermore, in each case, adding centre level mean KS2 scores into the regression improved the fit of the model (i.e., the predictions became more accurate).

In each subject, there was almost no difference between predictions from the prediction matrix method and the logistic regression without centre mean KS2 method. Therefore, we can attribute the changes to predictions using the logistic regression with centre mean KS2 method to the inclusion of centre mean KS2, rather than to the method itself.

The results suggest that using a logistic regression method should certainly be considered as part of the awarding process, as it would be lead to outcomes which are more closely aligned across specifications.

Perhaps more importantly, aside from the methodological recommendation, this report provides evidence that the current use of KS2 prediction matrices systematically leads to some AOs being lenient and others being harsh. Specifically, our evidence, combined with previous work in Benton & Sutch (2014), would suggest the model underpinning GCSE awarding has likely repeatedly underestimated how many students should achieve each grade or above for specifications with most of their entries from schools with high prior attainment. Conversely, for specifications with many entries from schools with low prior attainment, it has overestimated how many should achieve each grade or above. Furthermore, the scale of the differences in predictions can be substantially larger than the tolerances that are typically allowed in awarding.

Note that in 2026, as was the case in 2025, GCSE awarding will be based upon a “common centres” approach, where, broadly speaking, each AO simply maintains their own grade distribution with their own centres. However, the evidence in this report suggests that the standards AOs are carrying forward may be either too harsh or too lenient compared to others. We would recommend that this issue is investigated, and that necessary corrections are identified, ahead of awarding in summer 2026.

References

Benton, T. & Sutch, T. (2014). *Analysis of use of Key Stage 2 data in GCSE predictions.* Cambridge Assessment

Epple, D & Romano, R. E. (2011). Peer Effects in Education: A Survey of the Theory and Evidence. *Handbook of Social Economics*, 1, 1053–1163. Available at: https://doi.org/10.1016/B978-0-444-53707-2.00003-7

Hanushek, E. A., Kain, J. F., Markman, J. M. & Rivkin, S. G. (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics,* 18(5), pp 527-544. Available at: https://doi.org/10.1002/jae.741

Appendix A

The following figures show, for each subject, the percentage of candidates from each specification in each KS2 decile, where 1 is the highest attaining decile and 10 the lowest attaining. For example, Figure A1 shows that over 14% of candidates taking English literature spec 3 were in the highest attaining decile. Only just over 6% were in the lowest attaining decile. In all other specifications, the percentages in each decile were very close to 10%. Therefore, it is clear that spec 3 had higher attaining candidates than the other specifications.

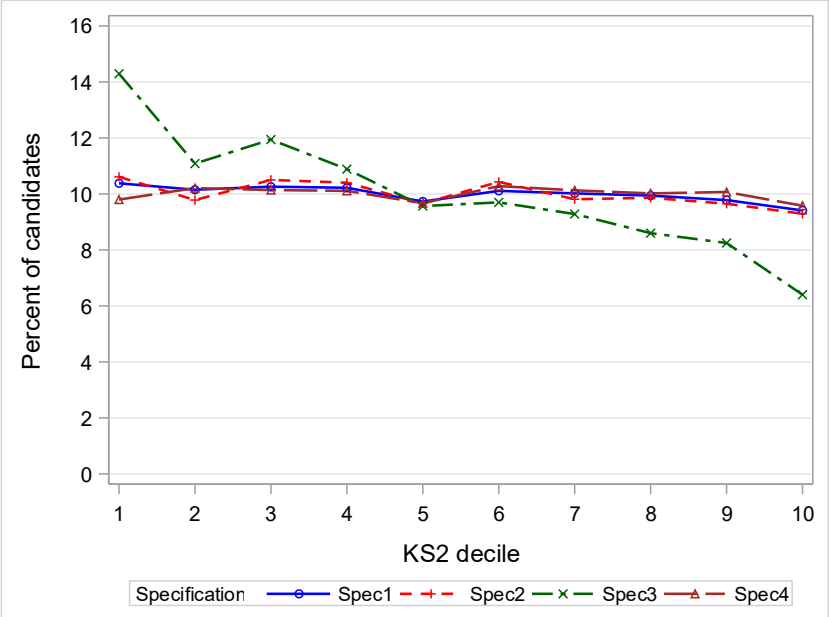

Figure A1: Percent of candidates in each KS2 decile, by specification (Eng lit.)

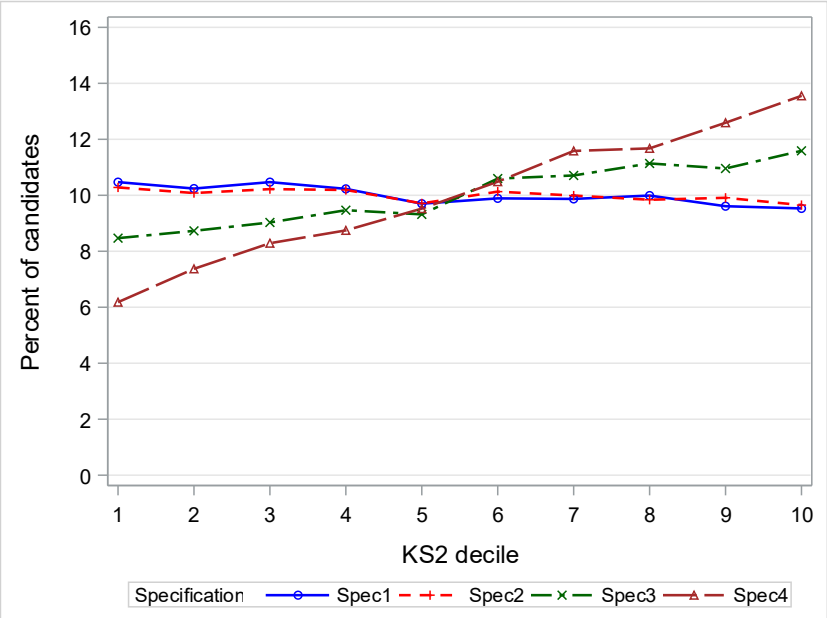

Figure A2: Percent of candidates in each KS2 decile, by specification (Maths)

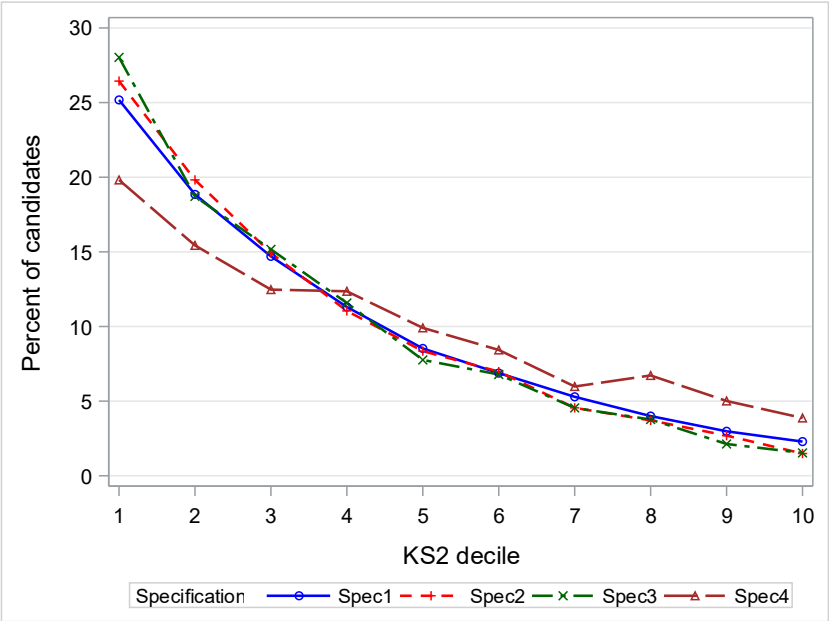

Figure A3: Percent of candidates in each KS2 decile, by specification (biology)

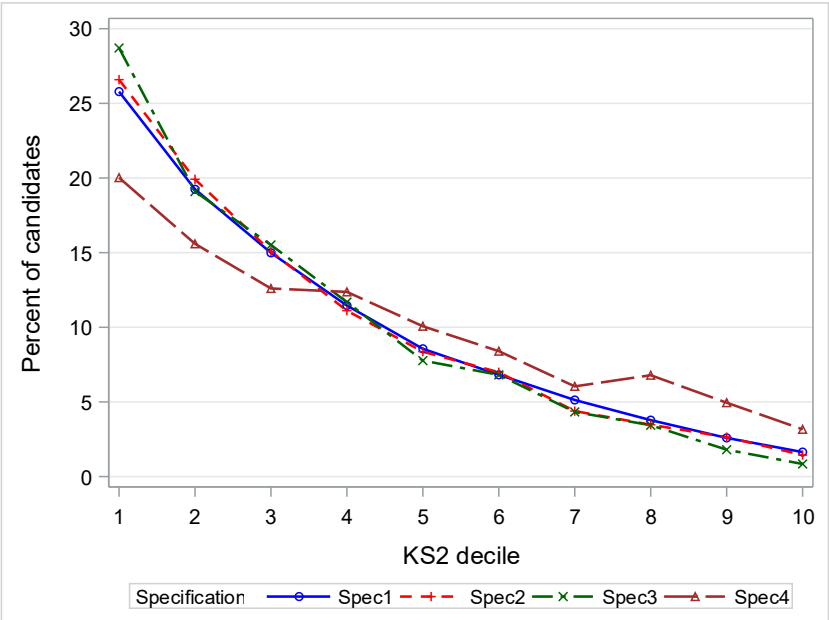

Figure A4: Percent of candidates in each KS2 decile, by specification (chemistry)

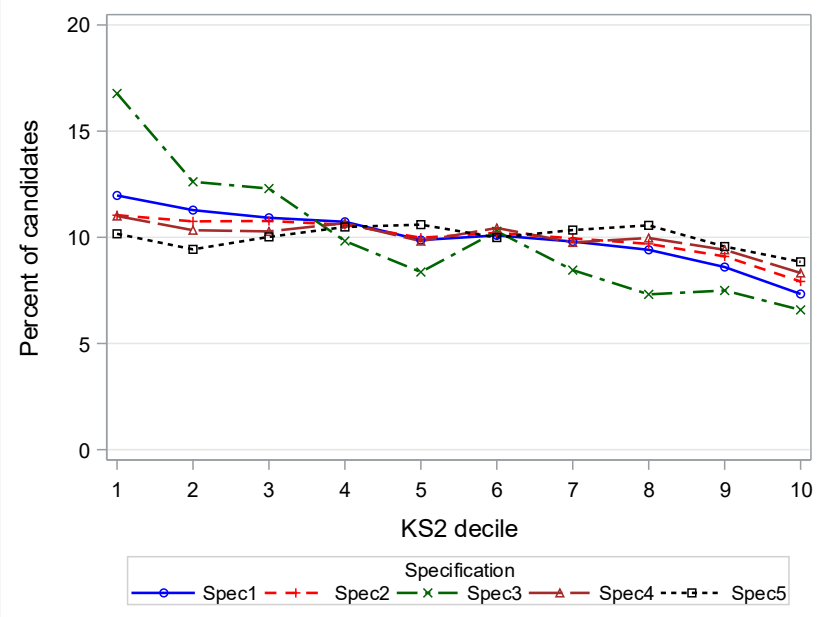

Figure A5: Percent of candidates in each KS2 decile, by specification (history)

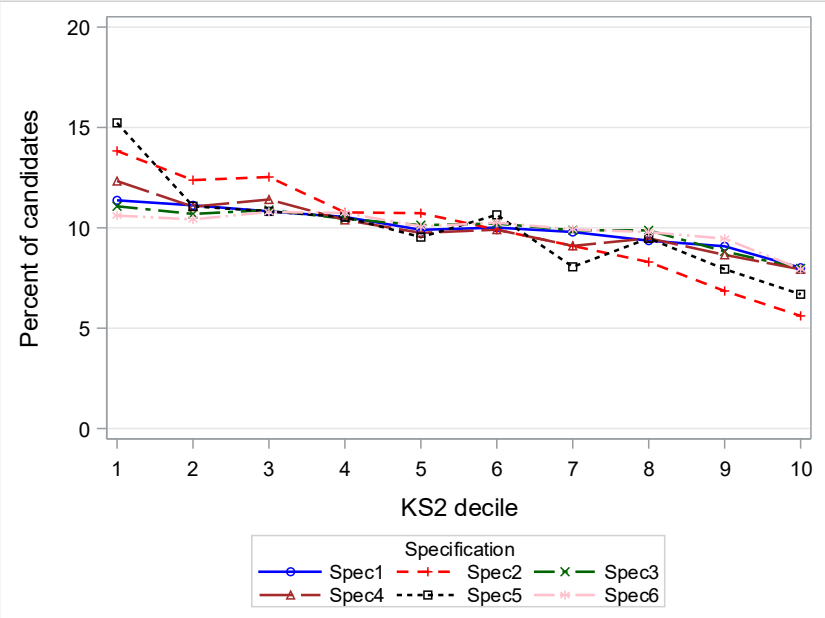

Figure A6: Percent of candidates in each KS2 decile, by specification (RS)

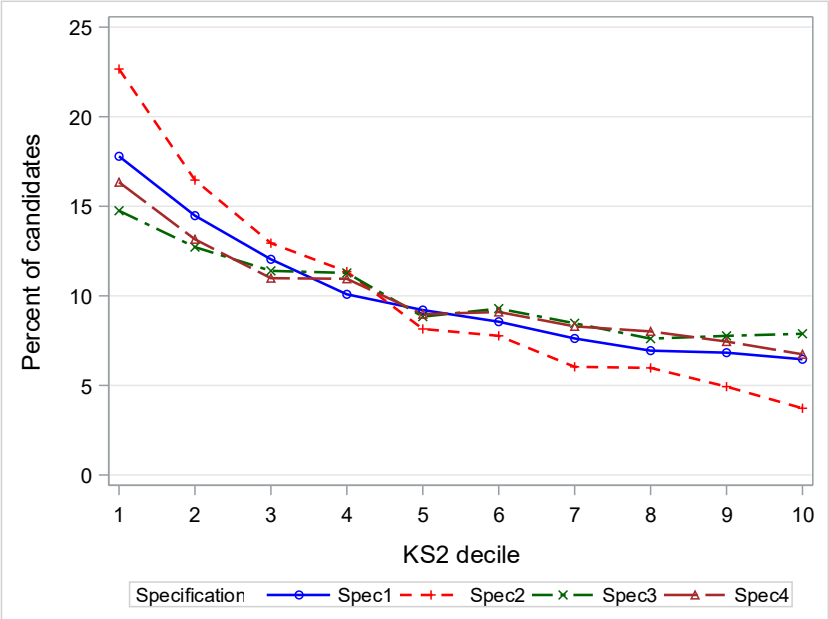

Figure A7: Percent of candidates in each KS2 decile, by specification (music)

Appendix B

In the logistic regressions, the standard errors were not adjusted for the clustering of students within schools. Technically, this may lead to SEs being underestimated, particularly for coefficients relating to the centre mean KS2. However, the SEs were all extremely small compared with the parameter estimates, so there was no chance that any of the effects would be non-significant had the adjustment been done.

Table B1: Parameter estimates for logistic regressions (model excluding KS2 centre mean, English literature)

Grade	Variable	Est	SE	P-value
9	Intercept	-8.333	0.041	<0.001
	KS2 score	0.081	0.001	<0.001
7	Intercept	-5.794	0.019	<0.001
	KS2 score	0.077	0.000	<0.001
4	Intercept	-2.518	0.013	<0.001
	KS2 score	0.076	0.000	<0.001
1	Intercept	1.399	0.027	<0.001
	KS2 score	0.062	0.001	<0.001

Table B2: Results of logistic regressions (model including KS2 centre mean, English literature)

Grade	Variable	Est	SE	P-value
9	Intercept	-11.142	0.106	<0.001
	KS2 score	0.077	0.001	<0.001
	Centre mean KS2	0.059	0.002	<0.001
7	Intercept	-8.534	0.052	<0.001
	KS2 score	0.074	0.000	<0.001
	Centre mean KS2	0.057	0.001	<0.001
4	Intercept	-5.269	0.045	<0.001
	KS2 score	0.073	0.000	<0.001
	Centre mean KS2	0.058	0.001	<0.001
1	Intercept	-1.749	0.121	<0.001
	KS2 score	0.057	0.001	<0.001
	Centre mean KS2	0.068	0.003	<0.001

Table B3: Parameter estimates for logistic regressions (model excluding KS2 centre mean, Maths)

Grade	Variable	Est	SE	P-value
9	Intercept	-10.632	0.049	<0.001
	KS2 score	0.115	0.001	<0.001
7	Intercept	-7.749	0.023	<0.001
	KS2 score	0.111	0.000	<0.001
4	Intercept	-4.125	0.015	<0.001
	KS2 score	0.113	0.000	<0.001
1	Intercept	0.490	0.026	<0.001
	KS2 score	0.100	0.001	<0.001

Table B4: Parameter estimates for logistic regressions (model including KS2 centre mean, Maths)

Grade	**Variable**	**Est**	**SE**	**P-value**
9	Intercept	-13.100	0.109	<0.001
	KS2 score	0.112	0.001	<0.001
	Centre mean KS2	0.052	0.002	<0.001
7	Intercept	-10.234	0.055	<0.001
	KS2 score	0.108	0.000	<0.001
	Centre mean KS2	0.052	0.001	<0.001
4	Intercept	-6.886	0.047	<0.001
	KS2 score	0.110	0.000	<0.001
	Centre mean KS2	0.058	0.001	<0.001
1	Intercept	-2.394	0.117	<0.001
	KS2 score	0.096	0.001	<0.001
	Centre mean KS2	0.063	0.002	<0.001

Table B5: Parameter estimates for logistic regressions (model excluding KS2 centre mean, biology)

Grade	Variable	Est	SE	P-value
9	Intercept	-8.104	0.062	<0.001
	KS2 score	0.086	0.001	<0.001
7	Intercept	-5.802	0.038	<0.001
	KS2 score	0.084	0.001	<0.001
4	Intercept	-4.325	0.047	<0.001
	KS2 score	0.121	0.001	<0.001
1	Intercept	1.727	0.103	<0.001
	KS2 score	0.063	0.002	<0.001

Table B6: Parameter estimates for logistic regressions (model including KS2 centre mean, biology)

Grade	Variable	Est	SE	P-value
9	Intercept	-11.690	0.133	<0.001
	KS2 score	0.081	0.001	<0.001
	Centre mean KS2	0.075	0.002	<0.001
7	Intercept	-9.564	0.088	<0.001
	KS2 score	0.079	0.001	<0.001
	Centre mean KS2	0.079	0.002	<0.001
4	Intercept	-8.495	0.128	<0.001
	KS2 score	0.115	0.001	<0.001
	Centre mean KS2	0.091	0.003	<0.001
1	Intercept	-1.446	0.304	<0.001
	KS2 score	0.053	0.002	<0.001
	Centre mean KS2	0.074	0.007	<0.001

Table B7: Parameter estimates for logistic regressions (model excluding KS2 centre mean, chemistry)

Grade	Variable	Est	SE	P-value
9	Intercept	-7.872	0.061	<0.001
	KS2 score	0.083	0.001	<0.001
7	Intercept	-5.443	0.037	<0.001
	KS2 score	0.078	0.001	<0.001
4	Intercept	-4.022	0.045	<0.001
	KS2 score	0.111	0.001	<0.001
1	Intercept	1.588	0.101	<0.001
	KS2 score	0.059	0.002	<0.001

Table B8: Parameter estimates for logistic regressions (model including KS2 centre mean, chemistry)

Grade	Variable	Est	SE	P-value
9	Intercept	-11.240	0.132	<0.001
	KS2 score	0.079	0.001	<0.001
	Centre mean KS2	0.071	0.002	<0.001
7	Intercept	-9.095	0.087	<0.001
	KS2 score	0.074	0.001	<0.001
	Centre mean KS2	0.077	0.002	<0.001
4	Intercept	-7.505	0.125	<0.001
	KS2 score	0.106	0.001	<0.001
	Centre mean KS2	0.075	0.002	<0.001
1	Intercept	-2.448	0.353	<0.001
	KS2 score	0.051	0.002	<0.001
	Centre mean KS2	0.089	0.008	<0.001

Table B9: Parameter estimates for logistic regressions (model excluding KS2 centre mean, history)

Grade	Variable	Est	SE	P-value
9	Intercept	-8.825	0.054	<0.001
	KS2 score	0.093	0.001	<0.001
7	Intercept	-6.121	0.026	<0.001
	KS2 score	0.087	0.000	<0.001
4	Intercept	-3.507	0.019	<0.001
	KS2 score	0.080	0.000	<0.001
1	Intercept	0.658	0.030	<0.001
	KS2 score	0.060	0.001	<0.001

Table B10: Parameter estimates for logistic regressions (model including KS2 centre mean, history)

Grade	Variable	Est	SE	P-value
9	Intercept	-11.596	0.129	<0.001
	KS2 score	0.090	0.001	<0.001
	Centre mean KS2	0.058	0.002	<0.001
7	Intercept	-9.008	0.070	<0.001
	KS2 score	0.084	0.000	<0.001
	Centre mean KS2	0.061	0.001	<0.001
4	Intercept	-6.337	0.059	<0.001
	KS2 score	0.077	0.000	<0.001
	Centre mean KS2	0.060	0.001	<0.001
1	Intercept	-2.516	0.131	<0.001
	KS2 score	0.056	0.001	<0.001
	Centre mean KS2	0.069	0.003	<0.001

Table B11: Parameter estimates for logistic regressions (model excluding KS2 centre mean, RS)

Grade	Variable	Est	SE	P-value
9	Intercept	-7.482	0.053	<0.001
	KS2 score	0.079	0.001	<0.001
7	Intercept	-5.111	0.028	<0.001
	KS2 score	0.074	0.000	<0.001
4	Intercept	-2.649	0.022	<0.001
	KS2 score	0.073	0.000	<0.001
1	Intercept	1.291	0.050	<0.001
	KS2 score	0.066	0.001	<0.001

Table B12: Parameter estimates for logistic regressions (model including KS2 centre mean, RS)

Grade	Variable	Est	SE	P-value
9	Intercept	-10.134	0.141	<0.001
	KS2 score	0.076	0.001	<0.001
	Centre mean KS2	0.055	0.003	<0.001
7	Intercept	-7.453	0.080	<0.001
	KS2 score	0.072	0.000	<0.001
	Centre mean KS2	0.048	0.002	<0.001
4	Intercept	-4.895	0.075	<0.001
	KS2 score	0.070	0.000	<0.001
	Centre mean KS2	0.047	0.001	<0.001
1	Intercept	-1.523	0.231	<0.001
	KS2 score	0.062	0.001	<0.001
	Centre mean KS2	0.060	0.005	<0.001

Table B13: Parameter estimates for logistic regressions (model excluding KS2 centre mean, music)

Grade	Variable	Est	SE	P-value
9	Intercept	-8.118	0.169	<0.001
	KS2 score	0.081	0.002	<0.001
7	Intercept	-5.604	0.087	<0.001
	KS2 score	0.075	0.001	<0.001
4	Intercept	-2.781	0.065	<0.001
	KS2 score	0.071	0.001	<0.001
1	Intercept	0.964	0.137	<0.001
	KS2 score	0.064	0.003	<0.001

Table B14: Parameter estimates for logistic regressions (model including KS2 centre mean, music)

Grade	Variable	Est	SE	P-value
9	Intercept	-12.318	0.410	<0.001
	KS2 score	0.076	0.002	<0.001
	Centre mean KS2	0.086	0.007	<0.001
7	Intercept	-9.505	0.242	<0.001
	KS2 score	0.071	0.001	<0.001
	Centre mean KS2	0.080	0.005	<0.001
4	Intercept	-6.640	0.233	<0.001
	KS2 score	0.067	0.001	<0.001
	Centre mean KS2	0.080	0.005	<0.001
1	Intercept	-3.197	0.604	<0.001
	KS2 score	0.058	0.003	<0.001
	Centre mean KS2	0.088	0.013	<0.001