Research and Analysis

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

Stephen Rhead, Qingping He, Beth Black and David Patchett from Ofqual's Strategy, Risk and Research Directorate

ofqual

Contents

Executive summary ... **3**

Introduction ... **4**

Data ... **5**

Analysis ... **7**

Centre based analysis ... *8*

Candidates based analysis ... *10*

Results ... **10**

Simple Analysis ... *10*

Centre based modelling ... *16*

Candidate based modelling ... *20*

Conclusion ... **31**

References ... **32**

Appendix ... **33**

Executive summary

Since 2013, Ofqual has looked at the pattern of variability in outcomes of schools and colleges (centres) over time for particular GCSE subjects as one way of understanding the extent of variability in the system. Overall, in years when specifications and overall cohorts are stable, one might expect the majority of centres with entries in successive years to have very similar outcomes.

In this report we are defining centre variability as the difference in proportion of candidates achieving grades A* to C across successive years within a subject at a particular centre. The majority of centres experience year-on-year stability. But of course where centres experience a drop in grade outcomes in one or more subjects, this can provoke speculation as to the possible causes. Some commentators have suggested that the comparable outcomes approach to awarding, in successfully managing unwarranted grade inflation, might be having a differential effect on those centres operating in a more challenging context.

In 2016 we published some initial analysis indicating centre variability was associated with centre type and the stability of the size of the entry, but was not associated with proportion of candidates entitled to free school meals, deprivation index of the centre (based on centre post code) or proportion of students with English as an additional language (Ofqual, 2016).

Other analysis, from Cambridge Assessment, has also looked to explain different levels of centre variability. Crawford and Benton (2017) concludes that most variation in centres' year-on-year outcomes is due to "well-known predictable-if-not-intuitive influence of chance in an indeterministic system" and also because of differences in the profile of student ability (as measured by concurrent mean GCSE). Essentially, the authors argue that 'volatility is normal'. This builds on previous research, also conducted by Cambridge Assessment, which indicated that even when marking reliability was maximised and, using simulations, grade boundaries chosen which minimise volatility, centre volatility would still remain (Bramley and Benton, 2015).

This report seeks to extend the analysis we presented in 2016 and existing research in this area. With access to additional data for those candidates sitting GCSEs in 2015 and 2016, including attainment at both GCSE and key stage 2, measures of socio-economic status (such as free school meals) and English as an additional language, we are better able to understand whether it is possible to predict if centres will experience stability or variability based upon the known profile of the centre cohort.

In this analysis, we looked at the influence of candidates' background characteristics on their performance in examinations and the variability of centre outcomes in successive years. We examined the effects using two approaches: analysis based on the average candidate profile of the centre, and analysis based on the profiles of individual candidates within the centre. We found that measures of socio-economic status have little or no bearing on centre variability. This indicates that the comparable outcomes approach to awarding does not have a systematic negative impact on centres with higher proportions of low socio-economic status candidates (similar to Ofqual, 2016).

It was found that attainment at either GCSE or key stage 2 is an important predictor of individual candidates' and individual centres' outcomes in any given year, although the attainment at GCSE is a better predictor. How the centre ability profile affects centre variability is not entirely clear although centres with very high or low ability profiles are more likely to experience lower level of variability in outcomes than centres with high proportions of grade C or grade D candidates. However, stable centres are not always those with low or high ability profiles.

Results from centre based analysis indicated that centres with a change in the number of students between years are more likely to experience variability in outcomes. Also, whether a centre was stable or variable in one year is a predictor of stability or variability in the following year. This means that centres who are stable in one year are likely to be stable the following year. However, centres who experienced positive volatility in one year are likely to experience negative volatility in the next. This is probably because such centres have a high proportion of candidates who are clustered around the grade C/D borderline (rather than, for example, clustering around grade B) and so more year-on-year variability is inevitable even if a small change in the entry size and/or ability profile.

Results from candidate based analysis indicated that the large majority of the observed variability in centre level outcomes can be attributed to changes in the ability profiles and other characteristics of successive years of candidates within centres. However, change in ability profile has the largest influence on centre variability. Only a small proportion of the observed centre variability may be associated with the indeterminacy of the examinations system.

Introduction

Since 2013, Ofqual has looked at the pattern of variability in outcomes of schools and colleges (centres) over time for particular GCSE subjects as one way of understanding the extent of variability in the system. Overall, in years when specifications and overall cohorts are stable, one might expect the majority of centres with entries in successive years to have very similar outcomes.

In recent years, Ofqual has published reports on the variability in GCSE results for centres with entry in both consecutive years (Ofqual, 2015), where centre variability was defined as the difference in proportion of candidates achieving grades A* to C across successive years. Most recently, Ofqual has also published interactive graphs[1] (Ofqual, 2017 and 2018) which allow users to explore the data by altering a number of features such as subject, size of entry and age group.

Indeed, the majority of centres display high levels of stability, i.e. there is little year-on-year variation. However, there are always centres which display large year-on-year variation. For individual centres, where there has been negative variability (i.e. an increase the proportion of lower grades) this is understandably worrying and probably is a catalyst for greater speculation and enquiry than when a centre experiences positive variability. Some commentators have suggested that the comparable outcomes approach to awarding, in successfully managing unwarranted grade inflation, might be having a differential effect on those centres operating in a

[1] https://analytics.ofqual.gov.uk/apps/2017/GCSE/CentreVariability/

more challenging context, for example, those with a significant percentage of students from low socio-economic status backgrounds or with speakers of English as an additional language (EAL). In 2016 we published some analysis indicating that centre variability was associated with centre type and the stability of the size of the entry, but was not associated with proportion of candidates entitled to free school meals (FSM), deprivation index of the centre, or proportion of students with EAL (Ofqual, 2016).

Other analysis, from Cambridge Assessment, has also looked to explain different levels of centre variability. Crawford and Benton (2017) concludes that most variation in centres' year-on-year outcomes is due to "well-known predictable-if-not-intuitive influence of chance in an indeterministic system" and also because of differences in the profile of student ability (as measured by mean GCSE grade across the subjects taken in the same year). Essentially, the authors argue that 'volatility is normal'. This builds on previous research also conducted by Cambridge Assessment which indicated that even when marking reliability was maximised and, using simulations, grade boundaries chosen which minimise volatility, centre volatility would still remain (Bramley and Benton, 2015).

This report seeks to extend the analysis presented in Ofqual, 2016, and other existing research in this area. With access to additional data for those candidates sitting GCSEs in 2015 and 2016, including attainment at key stage 2 (KS2) and GCSE, measures of socio-economic status (SES) (such as free school meals), gender, centre type, and EAL, we are better able to understand whether it is possible to predict if centres will experience stability or variability based upon known profiles of the centre cohorts.

Data

The data used in this study was obtained from the National Pupil Database (NPD) supplied by the Department for Education (DfE) and contained GCSE results for all Year 11 students (the target age group) in summer exam sessions between 2015 and 2016, prior attainment at KS2 and Schools Census data. To ensure the results are more meaningful, for each GCSE subject the proportion of candidates achieving grades A* to C was calculated for 2015 and 2016 for centres with at least 25 candidates in both years. The difference in the proportion of candidates achieving grades A* to C for 2016 compared to 2015 was calculated. As an example, the distribution of centre variability for GCSE mathematics is plotted in Figure 1. Each bar represents the number of centres with a particular level of variability given in intervals of two percentage points. Also reported are the mean centre variability and the standard deviation (SD) of the distribution. The mean centre variability is the average year-on-year difference for all of the centres, for example, a mean of 0.7% indicates that, on average, centres had a 0.7% increase in the proportion of candidates achieving a grade C or above, and the SD is a measure of the spread of the variability. The results for all subjects are summarised in Table 1.

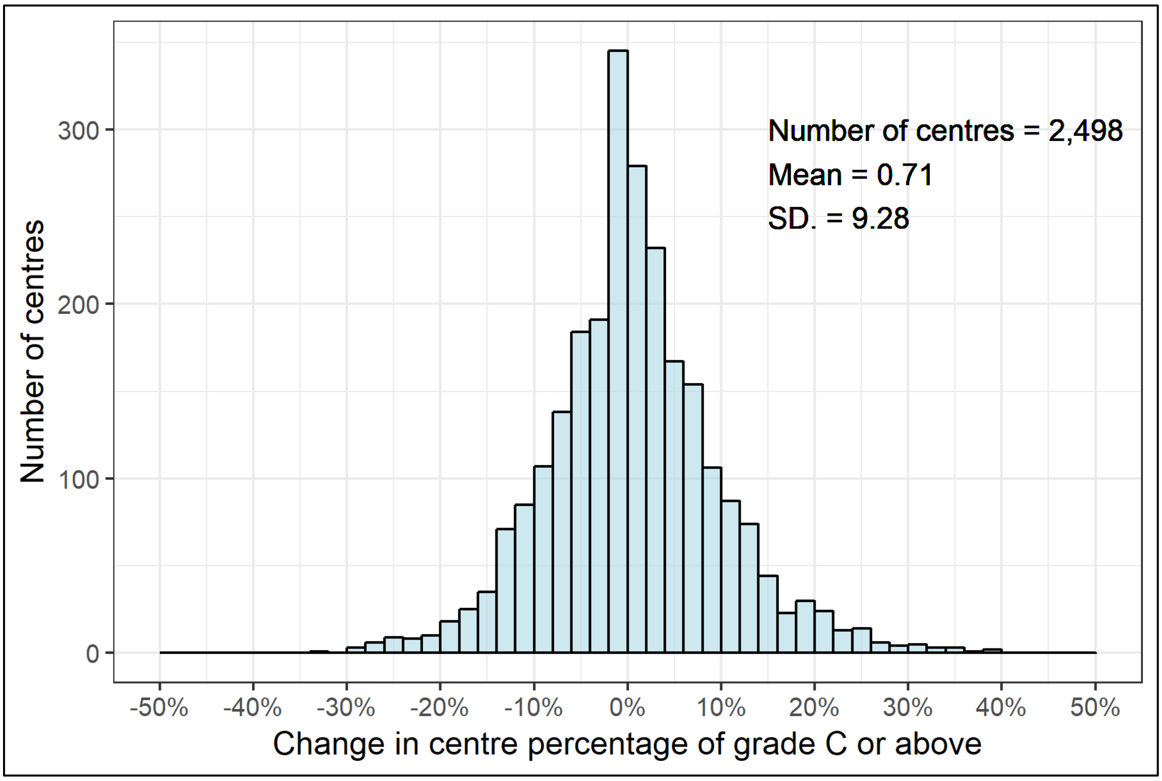

Figure 1. Difference in the proportion of candidates achieving grades A* to C in GCSE mathematics for summers 2015 and 2016.

Table 1. Difference in the proportion of candidates achieving grades A* to C in a selection of GCSE subjects for summers 2015 and 2016.

Subject	Number of centres	Mean centre variation (%A* to C)	SD. (%A* to C)
Mathematics	2,498	0.71	9.28
Additional science	2,032	-3.48	16.01
Science	900	-0.28	7.70
Biology	940	-0.52	8.70
Chemistry	968	-0.94	8.83
Physics	973	-3.61	18.01
English	1,362	-1.35	10.10
English literature	1,566	-1.49	12.40
Geography	1,610	-1.43	13.45
History	1,205	-2.24	11.98
French	1,029	-0.86	15.15
German	358	-1.07	13.58
Spanish	547	-1.12	14.74

Analysis

In the present study, we attempt to investigate the influences of candidates' background characteristics on the performance in examinations and the variability of centre outcomes in successive years using logistic regression and multilevel modelling. The following characteristics of the candidates have been explored:

- Ability as represented by concurrent attainment at GCSE or prior attainment at KS2;
- Gender;
- First language at home (English or non-English);
- Requirement for provision of special educational needs (SEN);
- Eligibility for free school meals (FSM);
- Level of deprivation as represented by the IDACI score[2];
- Type of centre (independent and selective vs others).

The attainment at GCSE of a candidate was represented using the mean of the GCSE grades in all GCSE subjects taken by the candidate in the same year, which involved converting GCSE letter grades to numerical values: A* to 8, A to 7, B to 6, etc. The prior attainment at KS2 was represented using the mean of KS2 mathematics fine level and English fine level. To get some sense of the potential influence of these factors on examination results, Table 2 shows the correlations between subject GCSE grade and the variables listed above for the 13 subjects in 2015 (the correlations for the 2016 dataset are shown in Table A1 in the appendix). When producing the table, the ability and the level of deprivation were treated as continuous variables, and the following were treated as binary variables: gender (0 for female and 1 for male), first language at home (0 for English and 1 for others), SEN status (1 for requiring SEN provision 0 and otherwise), eligibility for FSM (1 for eligible and 0 for illegible), and type of centre (1 for independent and selective centres and 0 for all other types of centre). As can be seen from Table 2, the correlation of subject grade with ability is strong and positive but weak with all the other variables, particularly with first language used at home. Girls generally had higher outcomes than boys. Candidates requiring SEN provision, who were eligible for FSM and with a high level of deprivation generally did not perform as well as other candidates in all subjects. Candidates in selective and independent centres had slightly better outcomes than candidates in other centres.

Table 2. Correlations of subject grade with background characteristics for the 13 subjects in 2015.

Subject	Mean GCSE	Mean KS2	First language	SEN	IDACI	FSM	Gender	Centre type
Maths.	0.90	0.58	0.03	-0.20	-0.12	-0.13	-0.08	0.07
Add. Sci.	0.90	0.53	0.03	-0.10	-0.13	-0.10	-0.11	0.11
Science	0.82	0.53	0.05	-0.12	-0.14	-0.11	-0.15	0.12
Biology	0.90	0.51	0.05	-0.09	-0.12	-0.09	-0.08	0.12

[2] IDACI is the income deprivation affecting children index and measures in a local area the proportion of children under the age of 16 that live in low income households. It is ranges from 0 (least deprivation) to 1 (highest deprivation).

Chemistry	0.86	0.67	-0.03	-0.29	-0.17	-0.17	-0.21	0.12
Physics	0.86	0.63	-0.01	-0.27	-0.17	-0.16	-0.21	0.12
English	0.91	0.65	-0.05	-0.21	-0.21	-0.17	-0.12	0.12
Eng. Lit.	0.82	0.55	0.07	-0.11	-0.10	-0.09	-0.14	0.12
Geog.	0.90	0.61	0.01	-0.19	-0.19	-0.16	-0.11	0.11
History	0.88	0.74	0.01	-0.28	-0.18	-0.17	0.02	0.12
French	0.89	0.54	0.03	-0.08	-0.13	-0.10	0.00	0.12
German	0.90	0.59	0.01	-0.26	-0.14	-0.15	-0.07	0.05
Spanish	0.80	0.52	0.07	-0.12	-0.13	-0.10	-0.14	0.12

We will first look at how the average profile of the candidates within a centre affects its variability of outcomes (centre based analysis) and then the influence of the profiles of individual candidates within the centre on centre variability (candidate based analysis).

Centre based analysis

In the first step of centre based analysis, each centre is assigned to one of the following three classifications of variability:

- Centres whose results have decreased year-on-year (negative unstable).
- Centres whose results are broadly similar to the previous year (stable).
- Centres whose results have increased year-on-year (positive unstable).

In addition, centres are classified as negative or positive unstable if there has been a large year-on-year decrease or increase in the proportion of candidates achieving A* to C. However, as there is no consensus as to what constitutes a large change, we choose a statistical measure of variation: the standard deviation[3]. The standard deviation is a measure used to quantify the amount of variation around the average of a distribution, where higher values indicate more variation. For each subject centres are classified based on the standard deviations reported in Table 1; centres are negative unstable if their variation is less than the mean centre variation minus the standard deviation, positive unstable if the variation is greater than the mean plus the standard deviation or stable otherwise. This process is illustrated by the shaded regions in Figure 2 and a breakdown by subject given in table A1 in the appendix.

[3] It is worth considering whether the standard deviation is appropriate to use as the stability threshold; by definition approximately 32% of centres are unstable. Does this lead to too broad a definition of stability, or should attention focus on more extreme variations (ie ± 25%)?

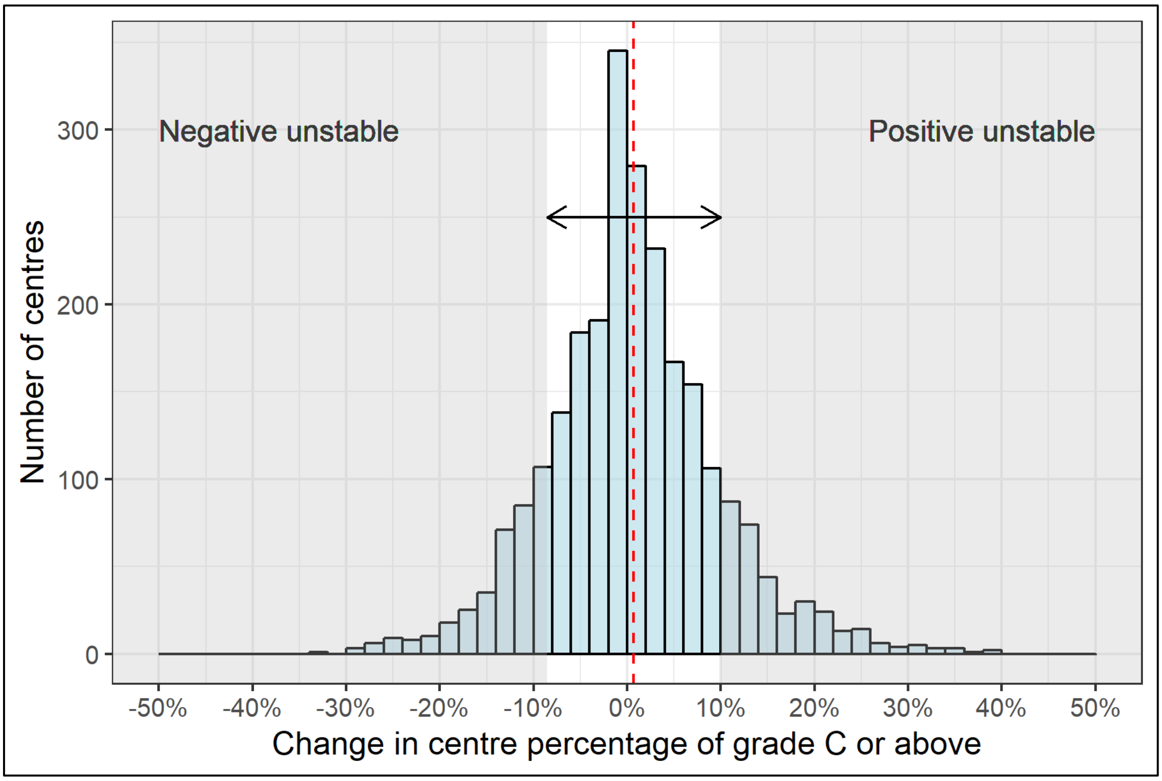

Figure 2. Difference in the proportion of candidates achieving grades A* to C in GCSE mathematics for summers 2015 and 2016. The mean centre variability is given by the vertical dashed line and the SD either side of the mean is denoted by the arrows. The shaded regions correspond to centres classified negative unstable and positive unstable.

The results of the classification for all subjects is tabulated in Table 3. We can see that if centres are classified stable in 2015 they are likely to be stable in 2016. Crucially there are very few centres that display large year-on-year decreases or increases in the proportion of candidates achieving a C or above; if a centre was classified as negative unstable in 2015 it is unlikely that the same centre in 2016 would expect further decreases in the proportion of candidates achieving a C or above, rather the centre would expect the proportion achieving a C or above to remain stable or to increase. The reverse is true for centres classified as positive unstable in 2015.

Table 3. Comparison of each centre's 2016 stability classification versus its 2015 stability classification.

	2016 Negative Unstable	*2016 Stable*	*2016 Positive Unstable*
2015 Negative Unstable	88	1,097	764
2015 Stable	1,353	9,507	1,106
2015 Positive Unstable	601	1,344	128

In the next step, we construct the average candidate profile using the set of characteristics discussed above for each centre for the 2015 and 2016 cohorts:

- The average ability measure of the centre represented using the average of the mean GCSE grade of all candidates in the centre or the average prior attainment at KS2 calculated as the mean of candidates' KS2 fine level
- The proportion of male/female candidates
- The proportion of candidates for whom English is an additional language
- The proportion of candidates with special educational needs provision
- The proportion of candidates eligible for FSMs
- The mean IDACI score of candidates

The proportion of candidates eligible for FSMs and the mean IDACI score were used as measures of socio-economic status. The changes between the 2015 and 2016 average candidate profiles were calculated as the yearly difference of these variables. We develop two centre based models to predict whether or not a centre can expect stability or variability based on the known profile of the cohort. In the first model we look at the average profile of the 2015 and 2016 cohorts (the static model) and in the second model we look at the change in average candidate profile (the dynamic model) using logistic regression. This will allow us to determine if absolute profile or the change in profile is more significant in predicting the stability of a centre. The development of these models is discussed in the results section.

Candidates based analysis

At individual candidate level, we examine how the performance of a candidate in an examination in a specific subject is affected by their background characteristics. More specifically, the probability of achieving a specific grade in a subject for each candidate in a particular cohort on an examination in a particular subject in a particular year is modelled using a statistical model – a two-level logistic regression model. For each centre, the model predicted probabilities of the candidates in the centre are aggregated to produce a predicted outcome for the centre which is then compared with the actual observed outcome of the centre to look at how the model performed. Changes in the centre level predicted outcomes between two years were then compared with the observed changes in outcomes to investigate the extent to which the observed centre variability can be explained by the model predicted variability. This approach is similar to that used by Crawford and Benton (2017).

Results

Simple Analysis

Having classified each centre as stable or unstable, for each subject, our attention turns towards building a model. The average profile of all candidates in a centre is calculated and we look to see if there are any similar characteristics within stable or unstable centres.

Initially we look at the average profile of the 2015 and 2016 cohorts (the static model) and then compare with the change in profile (the dynamic model). The distributions of the mean centre ability (represented by the centre mean GCSE grade) of the 2016 cohort may be seen in Figure 3 (the distribution of mean KS2 fine level is shown in Figure A1 in the appendix). Typically, stable centres are associated with a higher mean GCSE grade (suggesting that year-on-year there are fewer borderline grade C/D candidates than in unstable centres) and unstable negative centres are generally associated with a lower mean GCSE grade than other classifications. However, there is substantial overlap of the mean GCSE grade distribution for the three different classifications making it difficult to predict a centre's classification based solely on the mean GCSE grade of the 2015 and 2016 cohort alone. This is in contrast to the dynamic model where very clear associations with the changing mean GCSE grade and stability are observed (Figure 4). It is seen that if the mean GCSE grade of the cohort decreases between years it is likely that centres will see a drop in results. If the mean GCSE grade of the cohort increases centres results will likely improve. And if the mean GCSE grade is similar then results will be broadly similar. This seems to suggest that change in profile is a better predictor of stability/volatility than actual or absolute profile. For the prior attainment at KS2, the patterns are similar but are not as clear as those generated using mean GCSE grade (see Figure A2 in the appendix).

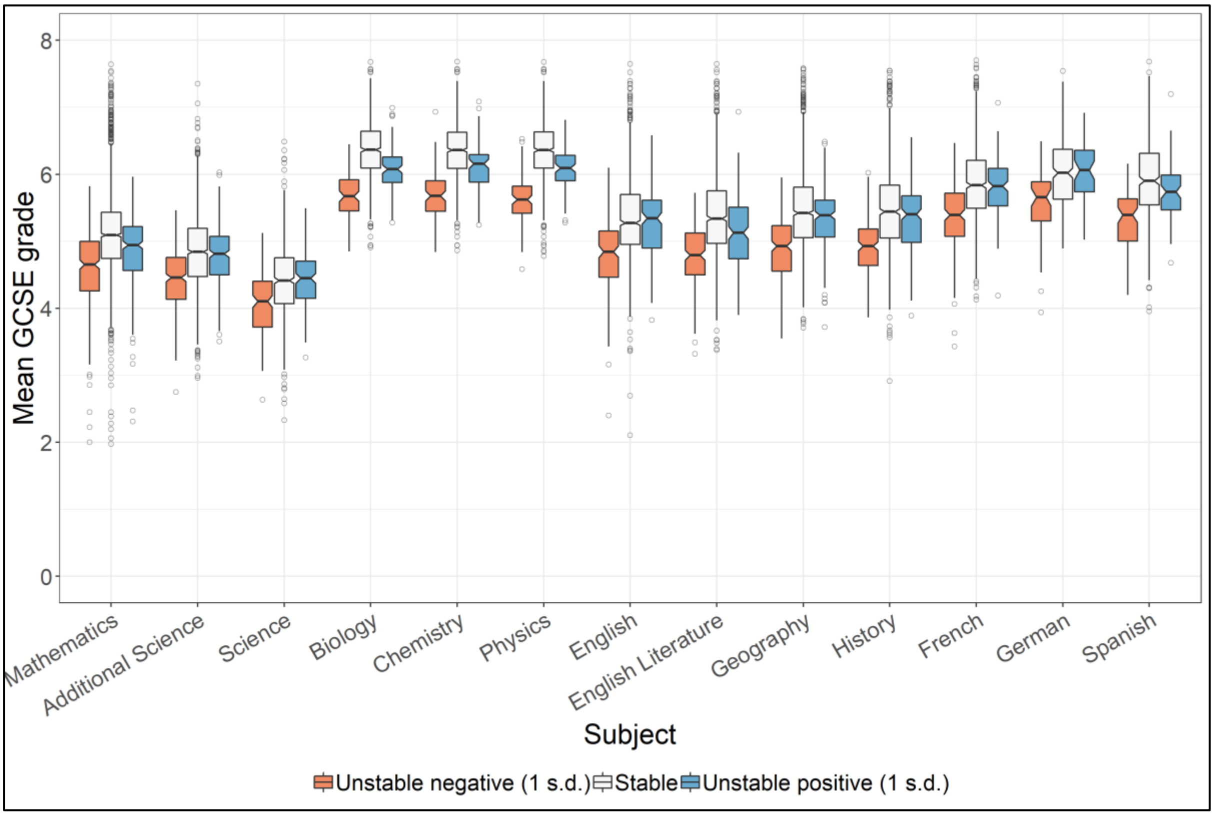

Figure 3. Distributions of centre mean GCSE grade for the 2016 cohort.

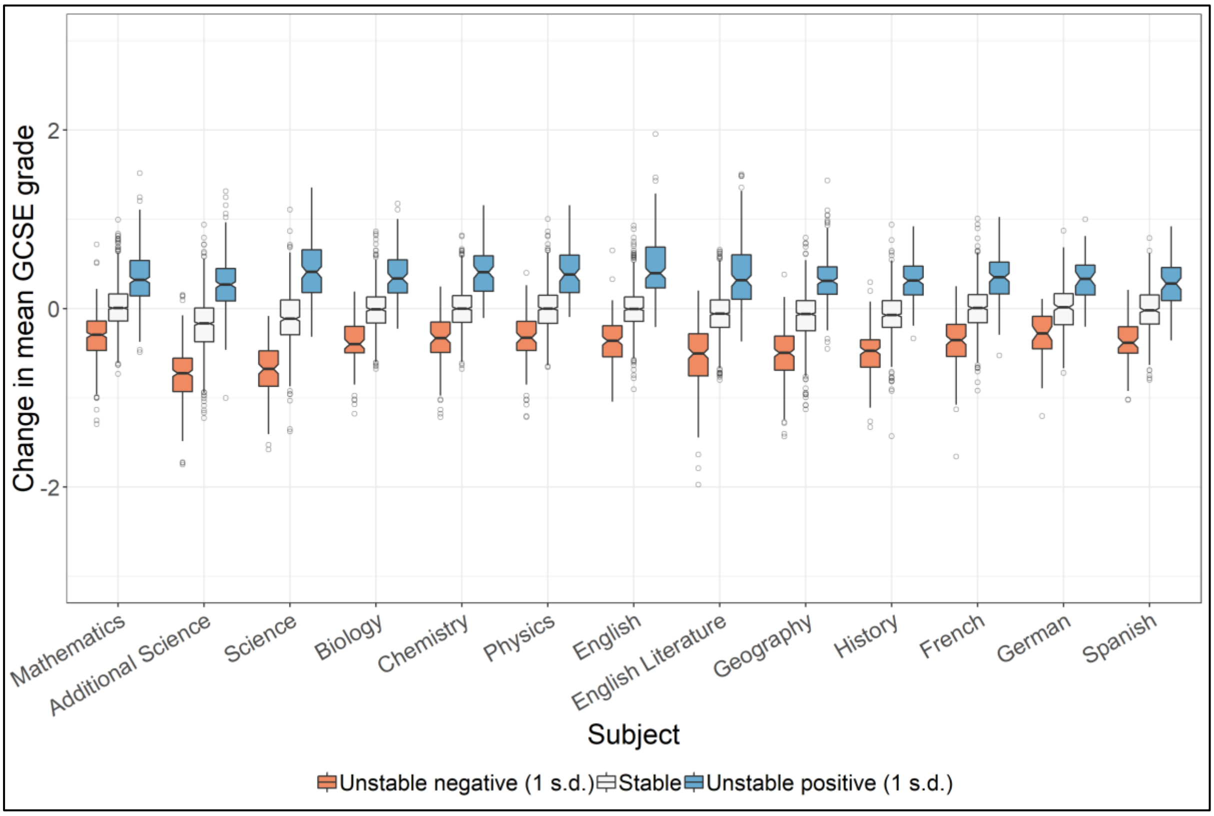

Figure 4. Distributions of changes in centre mean GCSE grade between the 2015 and 2016 cohorts.

Further comparisons between the absolute profile and change in profile may be seen in Figures 5 and 6, which illustrate the proportion of FSM eligibility of the 2016 cohort and the change in the proportion of FSM eligibility respectively. Unstable centres are associated with slightly higher levels of FSM eligibility; however, there is no clear relation between FSM eligibility and stability. There is some effect on the change in FSM eligibility and stability; unstable negative centres are associated with slight increases in FSM eligibility, FSM eligibility is broadly similar in stable centres and unstable positive centres are associated with slight decreases in FSM eligibility. However, it is worth noting that this association is far more moderate than changing mean GCSE grade.

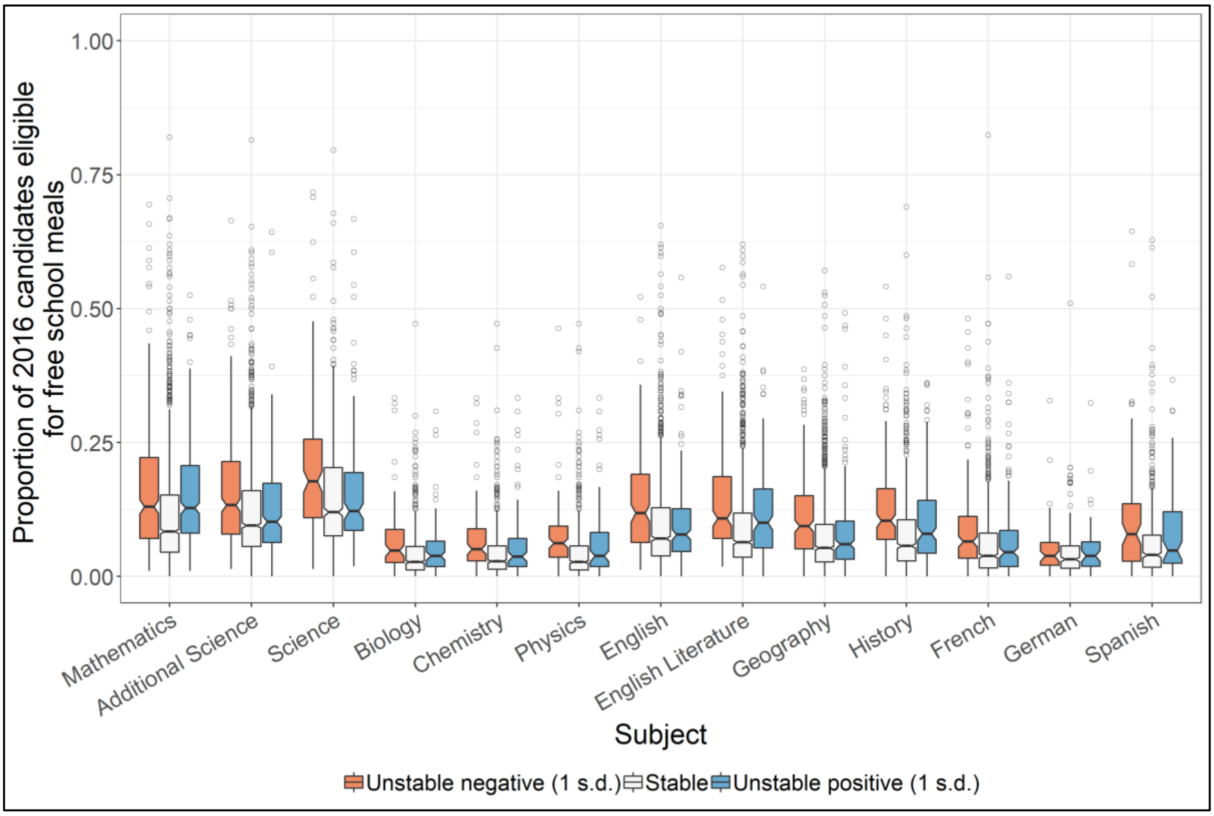

Figure 5. The proportion of FSM eligibility in the 2016 cohort.

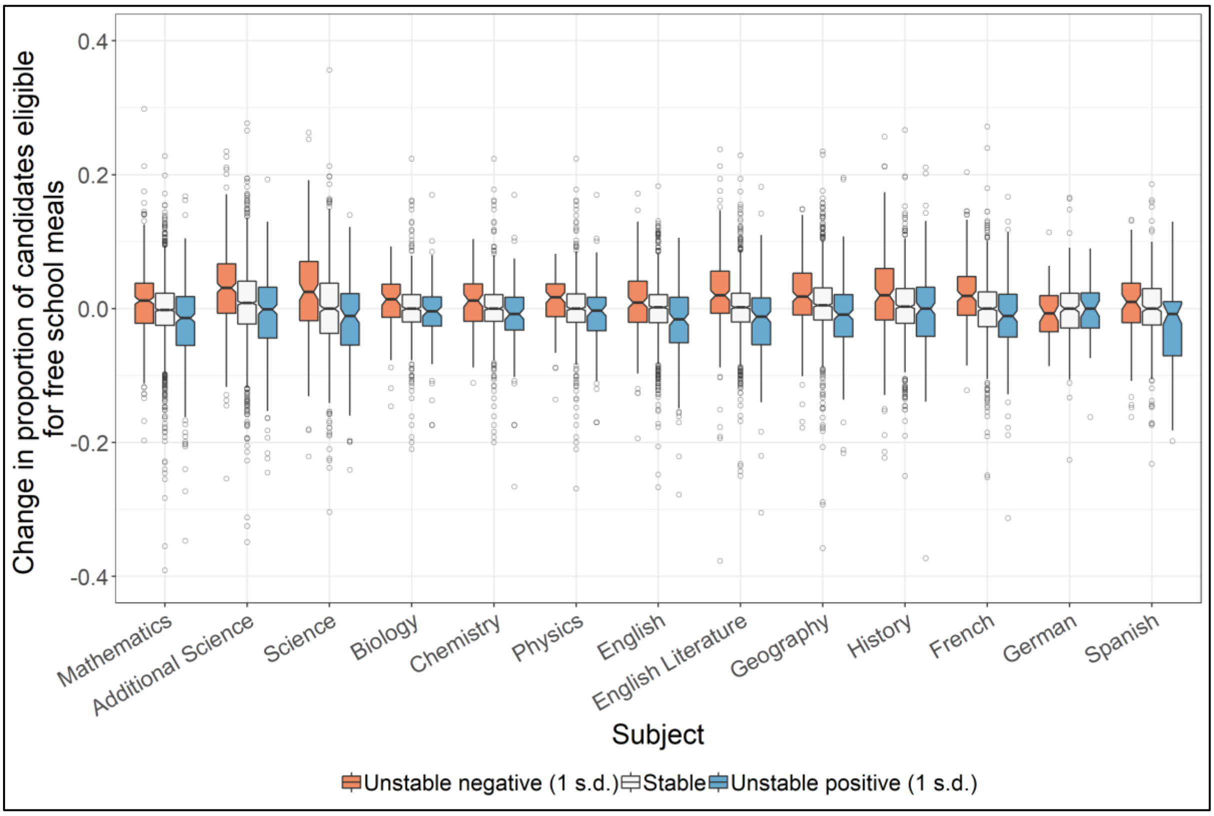

Figure 6. Change in the proportion of FSM eligibility.

Some associations are only observed for specific subjects which may reflect differences in the nature of subject and the associated examinations between the subjects. Figures 7 and 8 show the effect of the proportion of male candidates and the change in the proportion of male candidates with stability. There is little or no effect observed with the gender of 2016 cohort and stability, nor with the change in gender profile for most subjects. However, a strong effect is observed with change in gender profile and stability for English and English literature. Unstable negative centres are associated with an increase in the proportion of male candidates and unstable positive centres associated with a decrease in the proportion of male candidates.

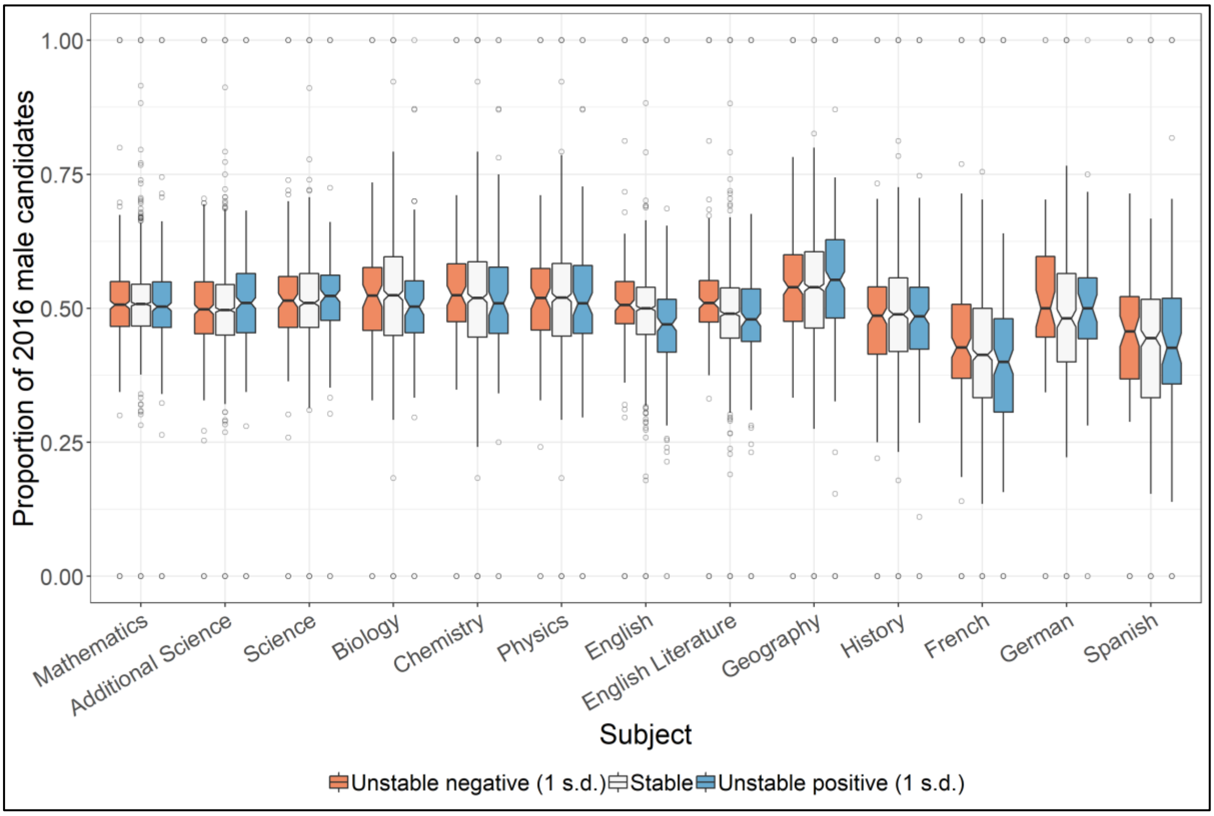

Figure 7. Proportion of male candidates in 2016 cohort.

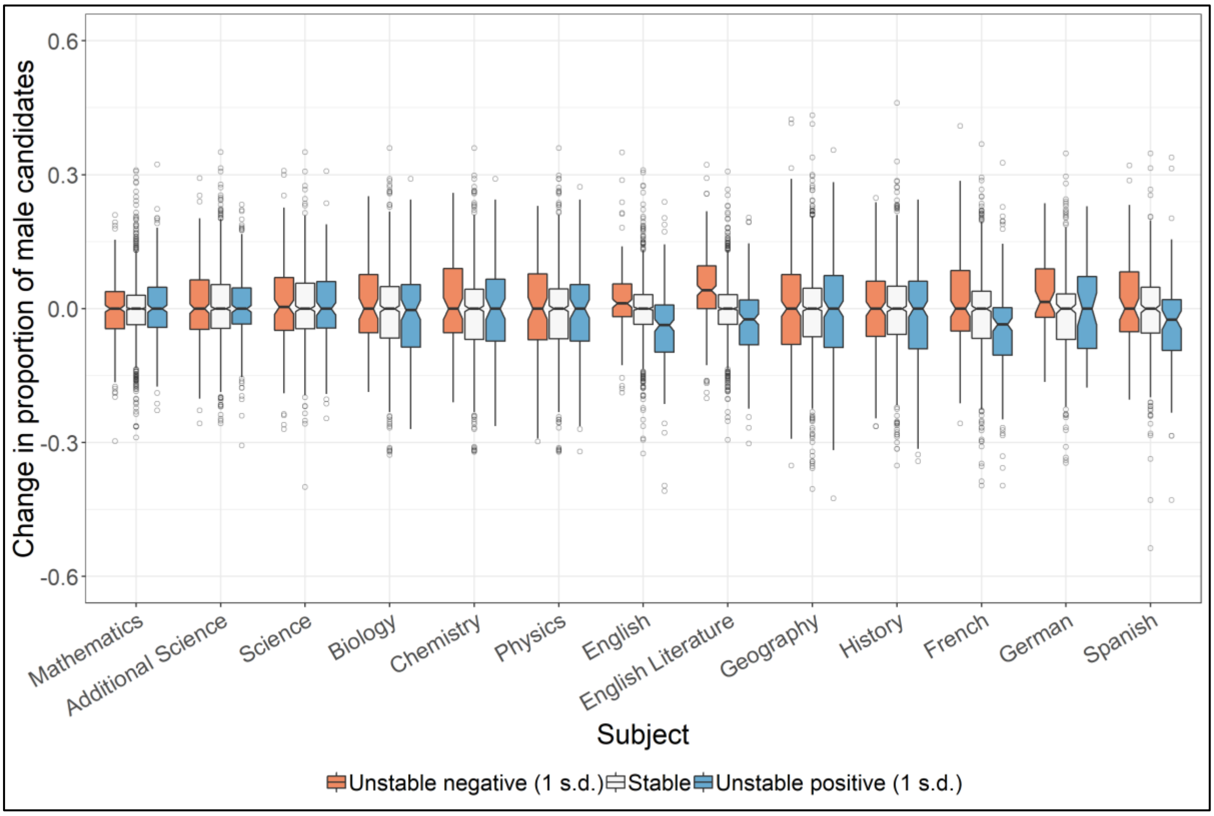

Figure 8. Change in proportion of male candidates.

The relationship between stability and various other measures of the average candidate profile may be seen in the appendix (Figures A3-A10) and the key findings summarised as:

- both types of unstable centres (positive and negative) appear to be associated with slightly higher IDACI scores
- there appears to be some association between a change in the profile of IDACI and stability
- there appears to be a discernible effect associated with the change in SEN profile and stability
- negative unstable centres are typically associated with larger percentage increases to cohort size whilst positive unstable centres are associated with smaller percentage increases.

Centre based modelling

Having conducted the simple analyses, it is necessary to combine the factors in a model which will help determine the extent to which any apparent effects are indeed significant.

The model for predicting the variability of centre outcomes

For each subject, a model was developed reflecting on the likelihood that the proportion of candidates in a centre achieving a grade C or above would be unstable. We focus on negative unstable and stable centres and introduce y_k as an indicator of stability classification for centre k, such that

$$y_k = \begin{cases} 1 \; for \; negative \; unstable \\ \quad 0 \; for \; stable \end{cases} \tag{1}$$

The developed model relates centre stability with the change in the centre average candidate profile and previous stability of the centre. These variables were chosen on the basis that they were likely to influence the stability as seen in Figures 5-8 and the graphs in the appendix). More specifically, the probability of a centre with a set of values on these variables being classified as a negative unstable centre is modelled using a logistic function, which is more conveniently given by the following equation:

$$\begin{aligned} logit(y_k) = \; & \beta_0 + \beta_1 . change \; in \; mean \; ability_k + \beta_2 . change \; in \; mean \; IDACI \; + \\ & \beta_3 . change \; in \; proportion \; of \; FSM \; eligibility \; + \\ & \beta_4 . change \; in \; proportion \; of \; SEN \; candidates \; + \\ & \beta_5 . change \; in \; proportion \; of \; non - native \; English \; speakers \; + \\ & \beta_6 . percentaage \; change \; in \; cohort \; size + \beta_7 . negative \; unstable 2015 + \\ & \beta_8 . positive \; unstable \; 2015 + \beta_9 . centre \; type \end{aligned} \tag{2}$$

where the logit $logit(y_k)$ is the logarithm of the odds ratio of centre k being classified as a negative unstable centre; β_0 to β_9 are model coefficients; negative (positive) unstable centres in 2015 were classified 0 if stable and 1 if negative (positive) unstable; and centre type takes 1 for selective and independent schools and 0 for all other types of schools.

Analysis based on attainment at GCSE

The model is first fitted using centre average GCSE grade as the ability measure and then average KS2 fine level as the ability measure. Results from the use of centre average mean GCSE grade are discussed below and those from average mean KS2 fine level are included in the appendix.

The significance of each factor influencing centre variability can be seen in Table 4. As is illustrated in Figure 4, change in the average ability of the candidates, as measured by change in average mean GCSE grade is a very strong predictor of centre stability. Unsurprisingly, centres who see large swings in pupil ability can expect year-on-year results to be unstable. Other factors that are consistently significant are whether or not the centre was classified as positive unstable in 2015. There are a few subject specific differences (such as the change in the proportion of boys being significant for English literature and history) but generally measures of deprivation are not significant. Centre type has little effect on stability once all other factors are taken into account. The model coefficients (β_0 to β_9) are not reported here as we are more interested in predicting the stability of the centres.

Table 4. Significance of each factor included in equation (2), broken down by subject (based on average centre mean GCSE grade).

Subject	Intercept	Centre type	FSM CHG	Male CHG	IDACI CHG	Ability CHG	Size CHG	Neg. 2015	Pos. 2015	SEN CHG	EAL CHG
Maths.	*					*		*	*		
Add. Sci.	*					*	*		*		
Science	*					*		*	*		
Biology	*					*			*		
Chemistry	*					*			*		
Physics	*					*	*	*	*		
English	*					*	*				
Eng. Lit.	*			*		*			*	*	
Geog.	*					*	*		*		
History	*			*		*			*		
French	*		*			*	*		*		
German	*					*			*		
Spanish	*					*			*		

The predicted probability of a centre being classified negative unstable is displayed by the boxplots in Figure 9. Due to the class imbalance of the data (approximately 14.6% of centres are negative unstable) the model itself is biased against centres classified as negative unstable, however we can see that the probability distributions for stable and negative unstable centres generally don’t overlap. The probability threshold of being classified negative unstable is calculated by maximising the Youden index (Bewick, Cheek, & Ball, 2004) and is given by the dashed black line in Figure 9. The percentage of centres classified stable or unstable are given in Table 5, and we can see that around 75% to 86% are correctly classified. This tells us that in a lot of instances variability is explainable and predictable based on known changes in the profile of candidates from one year to the next.

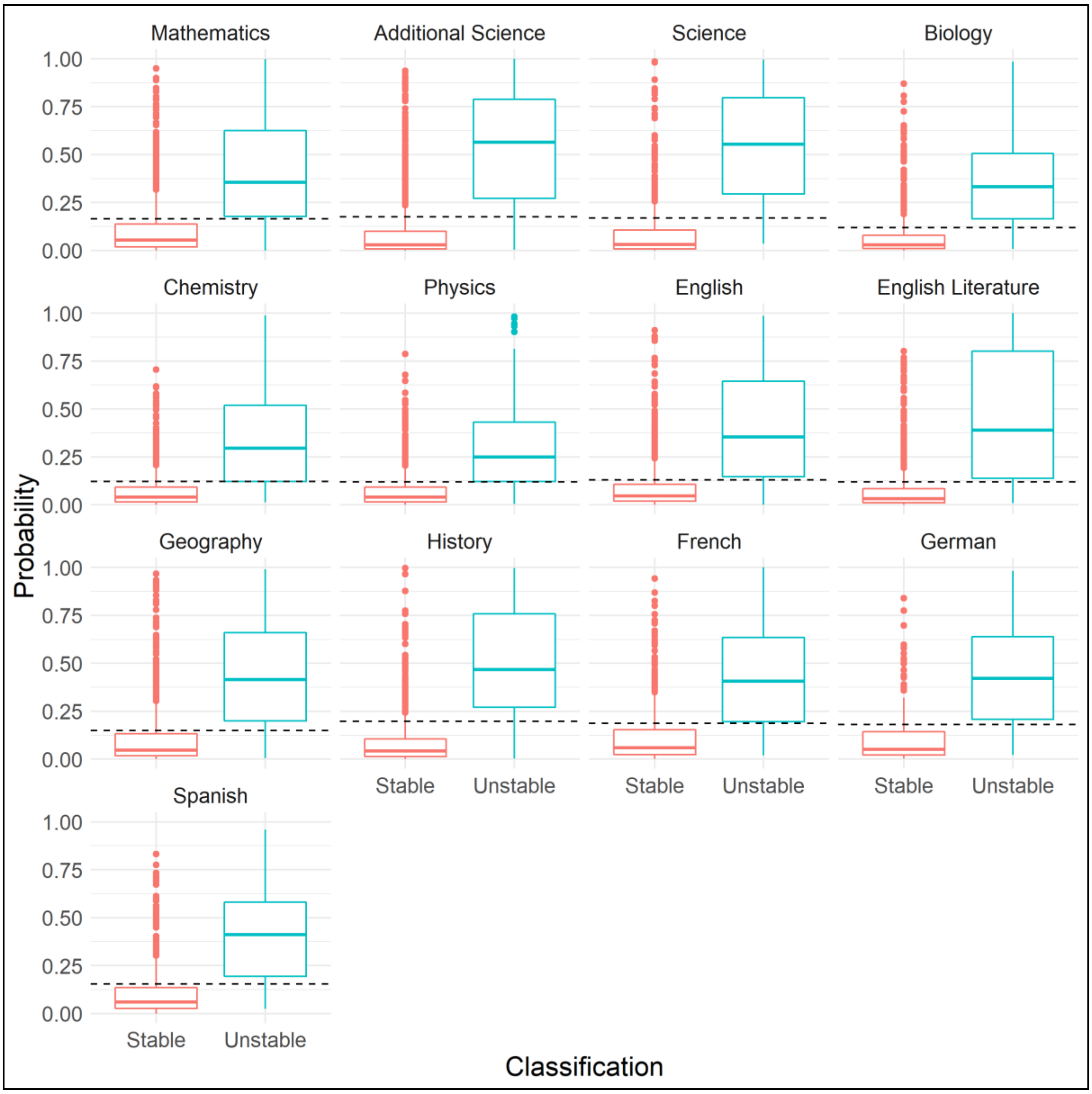

Figure 9. Probability of stable and negative unstable centres being classified negative unstable (based on centre average mean GCSE grade). The classification threshold is given by the black line.

Table 5. Percentage of centres classified stable or unstable correctly.

Subject	% correctly classified stable	% correctly classified negative unstable
Mathematics	78.6	77.5
Additional science	84.0	85.6
Science	83.5	83.6
Biology	83.4	78.6
Chemistry	81.8	75.2
Physics	80.8	76.3
English	78.8	78.0
English literature	81.4	79.1
Geography	77.9	83.4
History	86.3	84.0
French	79.7	78.1
German	82.0	74.5
Spanish	78.0	84.4

However, using this model a significant amount of variability is not explained. Other factors need to be considered as the model tends to classify negative unstable centres as stable if mean GCSE grade has increased compared to the previous year. There are other factors which cannot readily be factored into the model - for example, changes to teaching staff or approaches to teaching would likely affect centre variability but this cannot be modelled as data for this is not readily available. Furthermore, changes in centre outcomes may not be sensitively reflected by changes in centre level characteristics as centres with similar mean ability may have very different profiles for individual candidates. The candidate based modelling approach discussed below will provide further insight into the variability in centre outcomes.

Analysis based on attainment at KS2

The results of fitting the model with centre average KS2 fine level are shown in Table A3 and Figure A11 in the appendix. Although a considerable proportion of the centre variability can be explained by the model, the magnitude is substantially smaller.

Candidate based modelling

Having examined how centre level characteristics affect centre outcomes and changes in outcomes in successive years, we now turn our attention to the influence of the characteristics of individual candidates on centre outcomes and centre variability. This involves modelling the performance of individual candidates in exams based on their background characteristics and aggregating their predicted outcomes within centres to predict centre outcomes.

The model for predicting candidate's performance

Candidate level variables considered in the model are the same set of background characteristics used for centre based modelling. Let y_{ik} be the performance indictor of candidate i in centre k in an examination:

$$y_{ik} = \begin{cases} 1 \; if\ the\ the\ achieved\ grade\ is\ C\ or\ above \\ 0 \;\ if\ the\ the\ achieved\ grade\ is\ D\ or\ below \end{cases} \tag{3}$$

Similar to the centre based modelling, the probability p_{ik} of a candidate with a set of values in the background characteristics receiving a grade C or above is described using a two-level logistic function. With $logit(y_{ik}) = \log(p_{ik}/(1 - p_{ik}))$, the logarithm of the odds ratio, the model can be represented using the following linear equation:

$$\begin{aligned} logit(y_{ik}) = \ & \beta_0 + u_k + \ \beta_1.Ability_{ik} \ + \ \beta_2.Gender_{ik} + \ \beta_3.FSM_{ik} + \ \beta_4.IDACI_{ik} + \\ & \beta_5.Language_{ik} + \ \ \beta_6.SEN_{ik} + \beta_7.Year_{ik} + \beta_8.CentreType_{ik} + \\ & \beta_{9k}.CentreMeanAbility_k \end{aligned} \tag{4}$$

where:

- $Ability_{ik}$ = ability (mean GCSE grade or mean KS2 fine level, continuous);
- $Gender_{ik}$ = gender (binary, 0 for females and 1 for males);
- FSM_{ik} = eligibility for FSM (binary, 0 for illegible and 1 for eligible);
- $IDACI_{ik}$ = level of deprivation (continuous);
- $Language_{ik}$ = home first language indicator (binary, 0 for English and 1 for others)
- $Year_{ik}$ = year in which the examination was taken (binary, 0 for 2015 and 1 for 2016);
- $CentreType_k$ = type of centre k (binary, 1 for independent and selective schools and 0 for all other schools);
- $CentreMeanAbility_k$ = mean ability of centre k (average mean GCSE grade or average KS2 fine level, continuous);
- β_0 = intercept;
- β_1 to β_8 = model coefficients (fixed effect);
- u_k and β_{9k} = centre level random intercept and random coefficient.

The inclusion of centre mean ability as an independent variable is intended to take account of centre level effects such as the compositional effect. Initially the effect of centre level mean ability was treated as fixed. It was, however, finally treated as random as this improved the model fit substantially. The model was fitted for each of the subjects separately.

Analysis based on attainment at GCSE

Figure 10 provides a graphic representation of the estimates of the overall logistic regression coefficients, together with the estimated 95% confidence intervals, based on candidates' mean GCSE grade. It is to be noted that it is not appropriate to compare the relative importance of the factors in contributing to the predicted outcome as the scales of the variables are different. However, since ability has the largest scale (values vary from 0 to 8 in the case of mean GCSE grade and from 2 to 6 in the case of mean KS2 fine level), the magnitude of its coefficient suggests that it has the largest influence on exam performance.

For most of the variables and most of the subjects, the coefficients are significantly different from zero. School type is not significantly different from 0 except for mathematics for which candidates in independent and selective schools performed better than those from other types of schools. For eligibility for free school meals and level of deprivation, the coefficients for most of the subjects are not significantly different from 0, indicating that these two factors do not have differentiated influence on students' performance. For English, English literature, French, German and Spanish, girls generally performed better than boys. In contrast, boys performed better in mathematics and the sciences, particularly in physics and mathematics. As expected, candidates with non-English as the first language at home did not perform in English and English literature as well as those whose first language is English, but performed better in French, German and Spanish. Candidates performed similarly in 2015 and 2016 except for additional sciences for which candidates in 2016 performed slightly better. It is, however, noted that the data included in the analysis does not include all candidates taking the subjects in the two years.

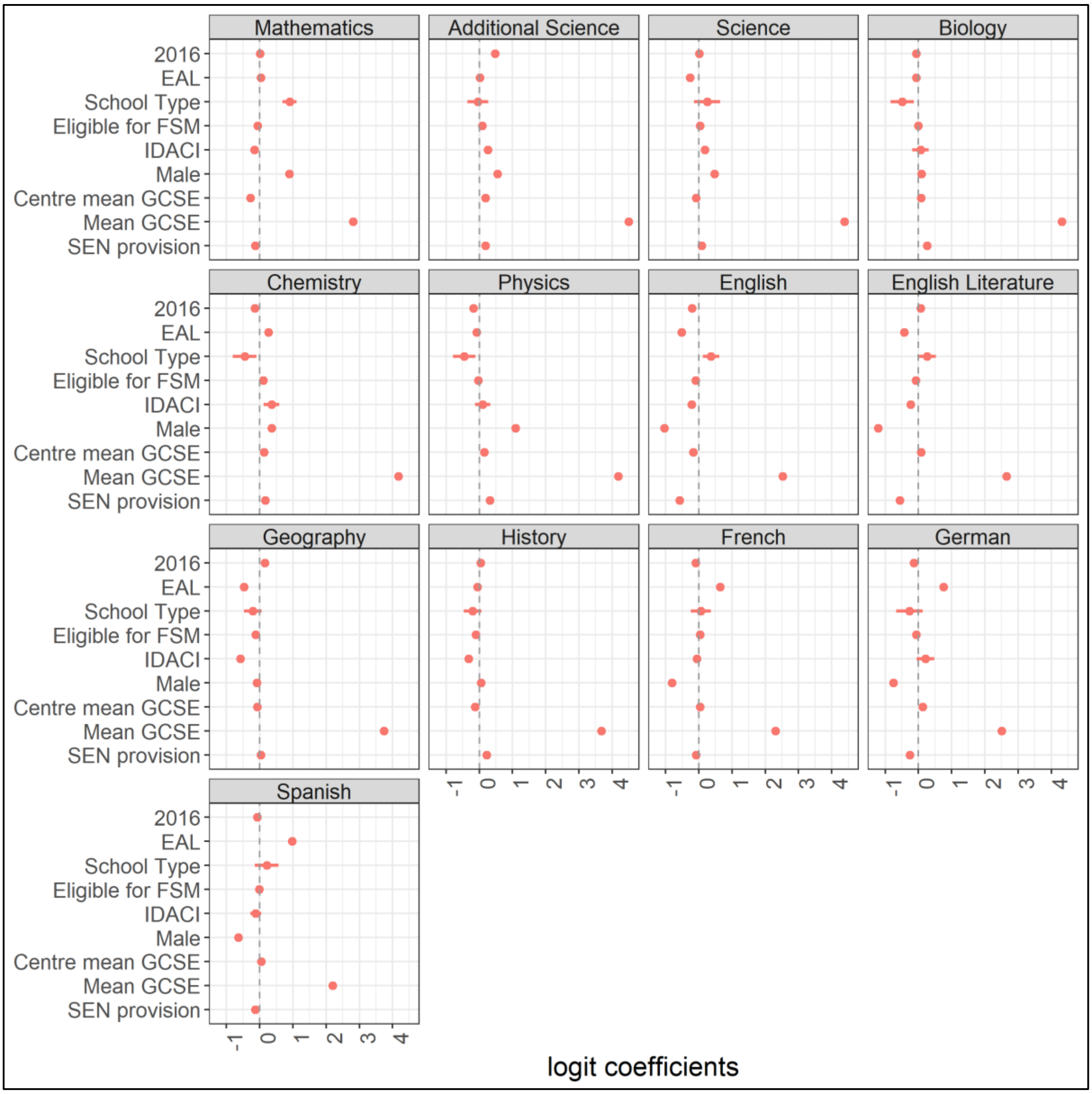

Figure 10. Distributions of the estimated model coefficients and the associated confidence intervals for the two-level logistic regression model using mean GCSE grade as the measure of candidate's ability for the 13 subjects studied.

To look at how the model fitted the data at individual candidate level, the box plots in Figure 11 compare the predicted probability distribution of candidates receiving a grade C or above based on their background characteristics for those who actually received a grade C or above and the probability distribution for those who only received a grade D or below for the 13 subjects. A perfect model-data fit would imply that those who actually received a C or above would have a predicted probability close to 1 while those who received a D or below a predicted probability close to 0. Figure 11 suggests that for all the 13 subjects, the predicted probabilities were over 0.80 for most of the candidates who actually were awarded a C or above. For those who received a D or below, although the probability distribution varies between the subjects, they were below 0.30 for most of the candidates. It is also worth noting that for all the subjects, the proportion of candidates receiving a C or above was substantially higher than that of candidates receiving a grade D or below, particularly for mathematics and the separate sciences. Overall, the model seemed to fit the data reasonably well in terms of its ability to predict.

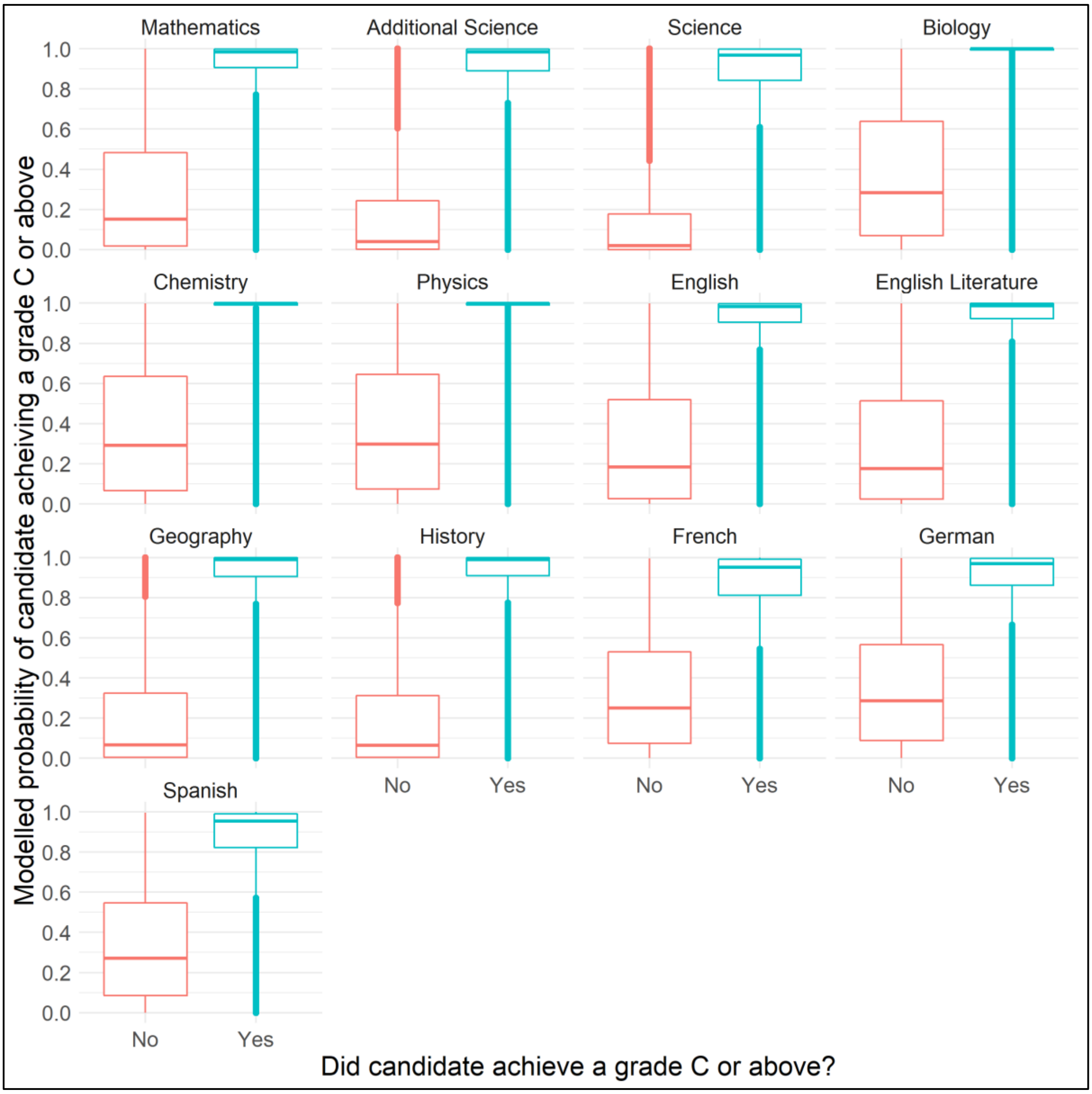

Figure 11. Distributions of model predicted probability for candidates receiving a grade C or above and the probability for those receiving a D or below based on their background characteristics for the 13 subjects studied (based on mean GCSE grade).

To see how well the model fitted the data at centre level, the predicted probabilities of the candidates within a centre is averaged (i.e. $P_k = \sum p_{ik}/n_k$, where n_k is the number of candidates in centre k) and then used as the model predicted proportion of candidates P_k achieving a grade C or above. Figure 12 compares the predicted proportions of candidates who achieved a grade C or above in centres and the actual observed proportions in 2015 and 2016. As can be seen, the modelled centre level proportions of C and above are highly consistent with the observed proportions for all the subjects in both years, with values of correlations between the predicted and observed proportions varying from 0.96 for German to 0.98 for English, suggesting that the model fitted the data very well at centre level.

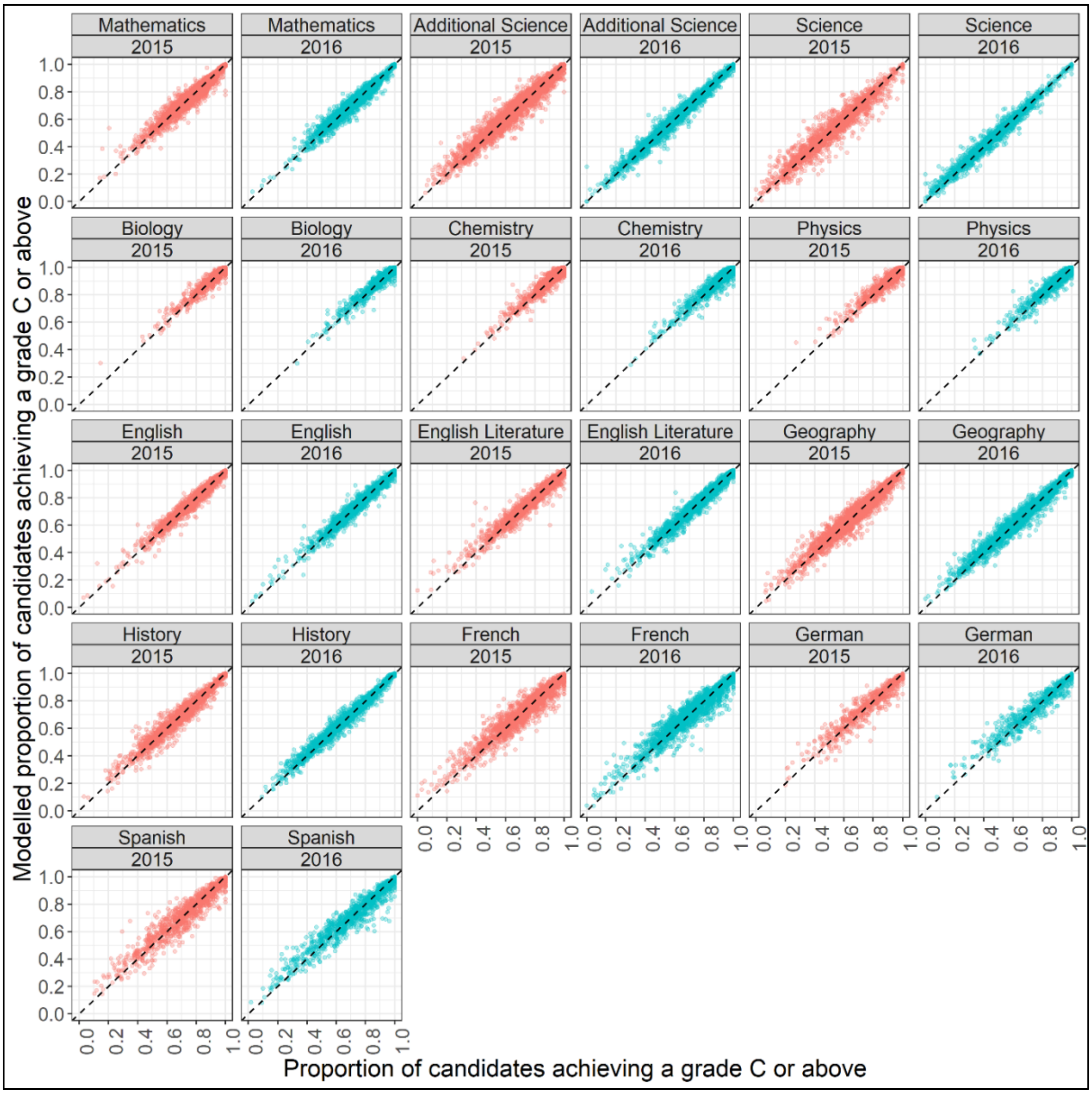

Figure 12. Comparison of modelled centre level proportion of candidates receiving a grade C or above and the observed proportion in 2015 and 2016 for the 13 subjects (based on mean GCSE grade).

The difference between the model predicted proportion of candidates receiving a grade C or above and the actual observed proportion for a centre will reflect the influence of other factors that are not included in the model. These factors may, for example, include the variability or uncertainty of performance of individual candidates in a particular exam and variability in the effectiveness of teaching and learning.

Figure 13 shows the distributions of the differences between the modelled proportions of candidates achieving a C or above and the observed proportions in centres in 2015. For mathematics and the separate sciences, the differences for most of the centres are within ±5pp. They are slightly larger for additional science, science, French, German and Spanish, but are still generally within ±10pp. As is clear, these distributions are considerably narrower than the distributions of the differences in the observed proportions in centre outcomes between two consecutive years that are normally seen. Differences in the distributions between the subjects will to a large extent reflect the difference in the nature of subject and the associated examinations. These distributions would be similar to the expected distributions of changes in centre proportions of candidates with a C or above between 2016 and 2015 if the profiles of the candidates (as represented by numbers of candidates and the values on the various characteristics included in the model) in centres, the effectiveness of teaching and learning, as well as the exams were to remain the same in 2016.

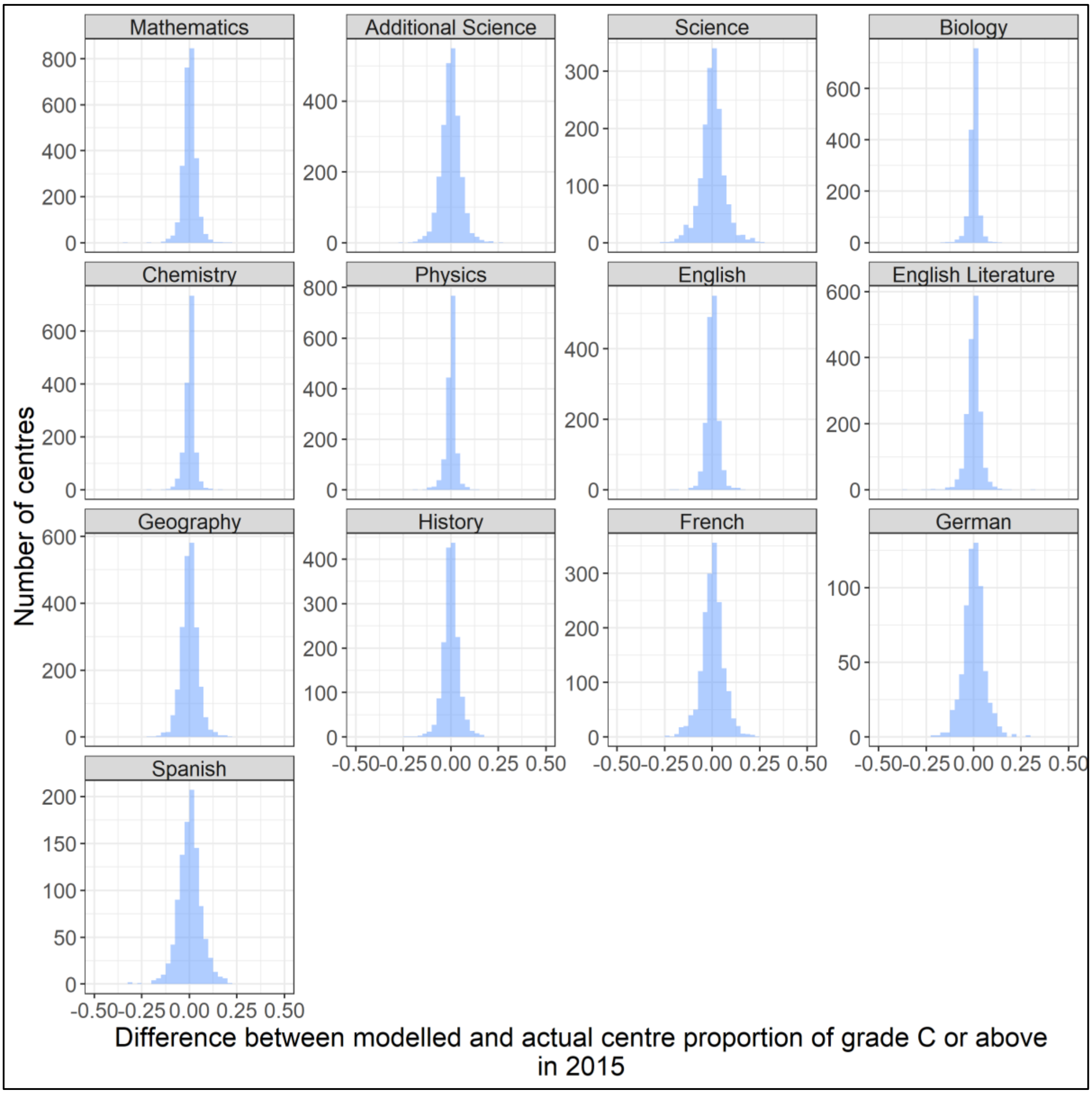

Figure 13. Frequency distribution of the difference between observed centre level proportion of candidates achieving a C or above and the model predicted proportion for the 13 subjects in 2015 (based on mean GCSE grade).

As has been demonstrated, the characteristics of candidates included in the model can exert various degrees of influence on their performance in exams. Changes in these characteristics and the entry size as well as the examinations over time will produce additional contribution to the variability associated with factors not considered (e.g., the graphs shown in Figure 13). To see this, Figure 14 compares the distribution of the differences of the observed centre proportions of grade C or above between 2016 and 2015 and that of the differences of the modelled proportions (i.e. $P_{k,2016} - P_{k,2015}$) for the 13 subjects. The correlations between the observed changes and the model predicted changes between the two years varied from 0.80 for Spanish to 0.94 for science (see Table 6), suggesting that the majority of the observed variability in centre outcomes between 2016 and 2015 in the 13 subjects can be attributed to the variability in modelled proportions as a result of changes in candidates' profiles between the two years. The proportion of the observed variability in centre outcomes between 2016 and 2015 that can be explained by changes in candidates' profile varies from 64% for Spanish to about 89% for science. Since candidate's ability has the largest influence on exam performance, centres that experience larger changes in candidates' profiles in terms of ability distribution may see large changes in outcomes (also see previous discussion). As discussed earlier, when the abilities of successive cohorts of candidates in a centre increase, its outcomes are likely to improve which will result in a positive change in its outcome. Similarly, if the abilities of successive cohorts of students decrease, the outcomes are likely to decrease. Table 6 also shows the correlation between changes in centre level proportions of C and above from 2015 to 2016 and the corresponding changes in centre level mean GCSE grade, which are positive and moderately strong for the majority of the subjects. Of course, the actual impact of the change of candidates' ability profile is difficult to assess as the modelling is undertaken at individual candidate level.

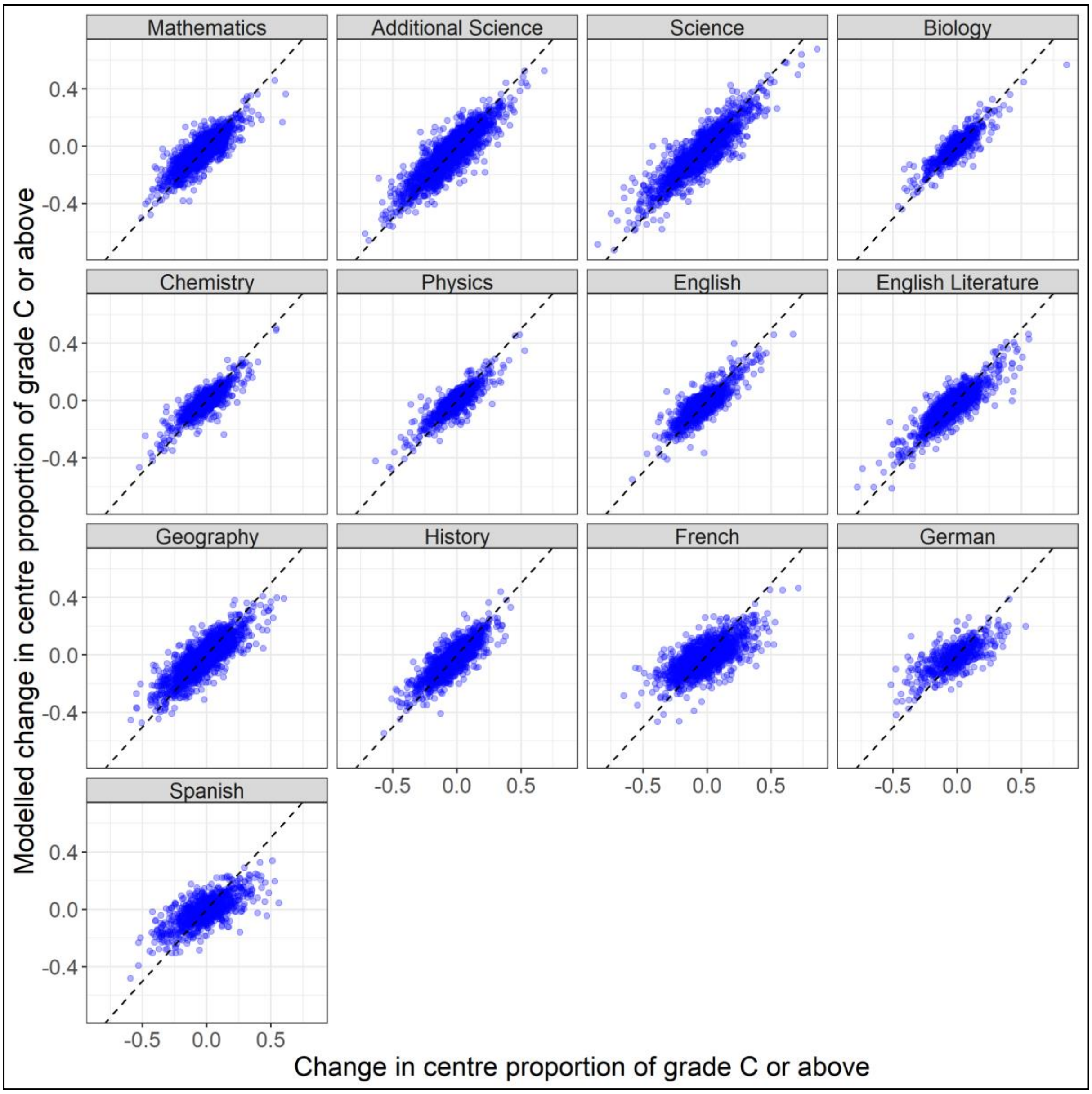

Figure 14. Comparison of observed centre level variability in outcomes between 2016 and 2015 and model predicted variability (based on mean GCSE grade).

Table 6. Correlations between model predicted changes in centre outcomes and the observed changes, and between observed changes and changes in centre average mean GCSE grade.

Subject	Correlation coefficient modelled changes in centre outcomes vs observed changes	Correlation coefficient observed changes centre outcomes vs changes in average mean GCSE grade
Maths.	0.94	0.54
Add. Sci.	0.89	0.27
Science	0.83	0.27
Biology	0.87	0.29
Chemistry	0.89	0.45
Physics	0.92	0.42
English	0.88	0.44
Eng. Lit.	0.79	0.29
Geog.	0.87	0.48
History	0.90	0.68
French	0.89	0.32
German	0.94	0.46
Spanish	0.80	0.23

As will be seen later, entry size and its change over time will have important impacts on the variability of centre outcomes. In particular, large changes in entry size for small centres may result in large changes in outcomes.

Figure 15 further compares the frequency distributions of the modelled changes in proportions of candidates with a grade C or above (blue bars) and the frequency distributions of the changes in observed proportions (red bars) between 2016 and 2015 for the 13 subjects. These distributions reflect the relationships shown in Figure 14. The graphs show that modelled variability is smaller than the actual observed variability. As discussed earlier, even if the candidates' profiles remained the same, a certain degree of variability would still exist as a result of uncertainty associated with the performance of candidates in examinations. The modelled proportions of C and above in centres and the differences of the modelled proportions between years will therefore have errors or uncertainties. We have adopted the procedure used by Carol and Benton (2017) to estimate the errors associated with estimated or modelled centre proportion C or above and the modelled difference of proportions between 2016 and 2015 for individual centres. The variance associated with a centre estimated proportion P_k is a function of the probability distribution of the candidates in the centre and its entry size (the smaller the entry size, the larger the uncertainty) and calculated as $\sigma_k^2 = \sum p_{ik}(1-p_{ik})/n_k$. The variance associated with the difference of the modelled centre proportion between two years $P_{k,2016} - P_{k,2015}$ is calculated as the sum of the variance in the respective year, i.e. $\sigma_{k,2016-2015}^2 = \sigma_{k,2015}^2 + \sigma_{k,2016}^2$. These variances associated with the centre modelled differences were then used to calculate the probability distribution against the level of variability of in centre outcomes of all the centres between 2016 and 2015. This probability distribution is superimposed on the frequency distributions of the modelled and observed differences in centres outcomes between 2016 and 2015 and represented by the smoothed line in Figure 15. The model predicted probability distribution and the actual frequency distribution are very close, suggesting once again that the model fits the data very well.

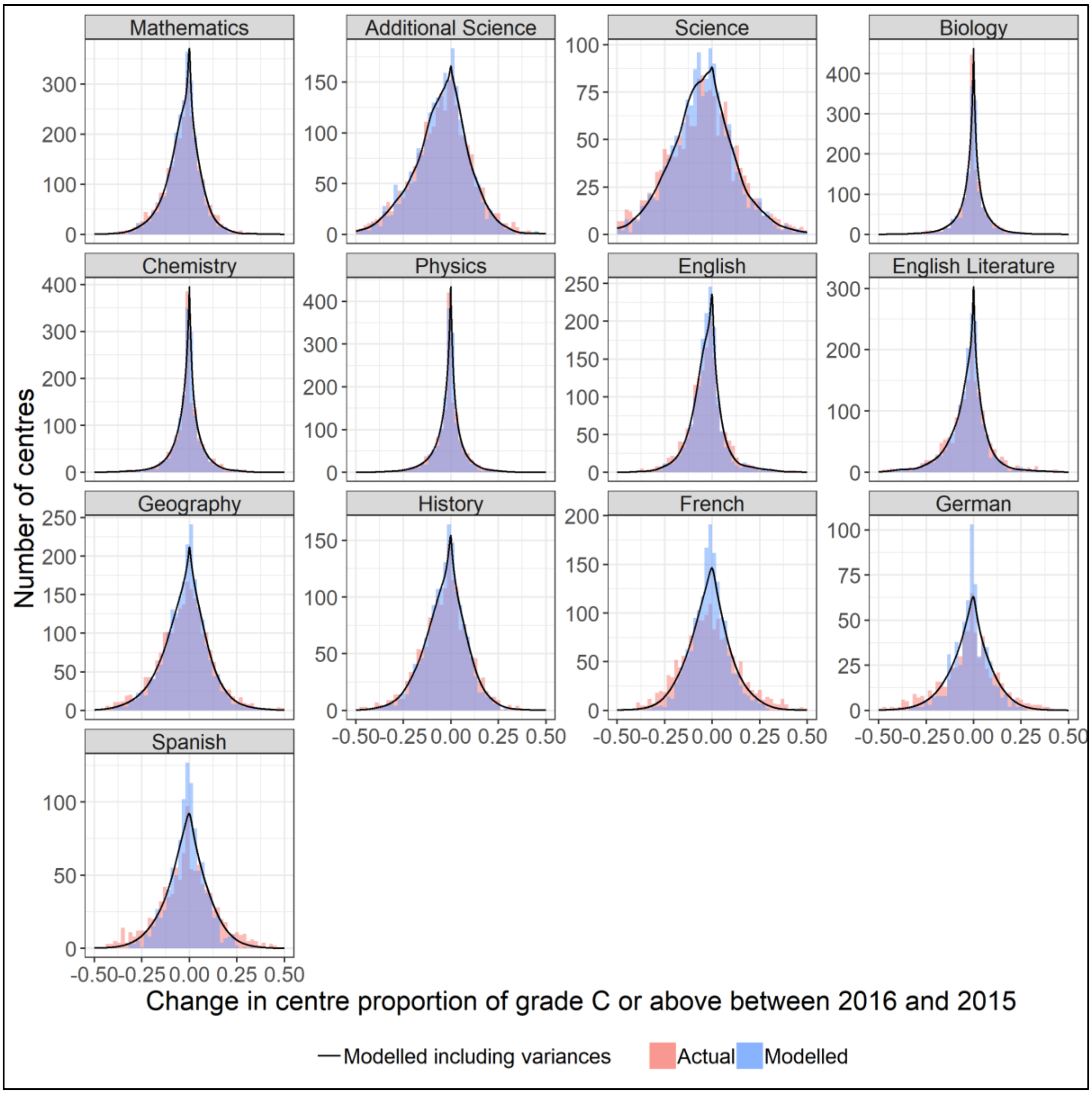

Figure 15. Comparison of the frequency distributions of the observed centre level changes in proportions of candidates achieving C or above between 2016 and 2015 (red bars) and the model predicted changes (blue bars) and the modelled probability distributions against the level of changes for all centres (the smoothed black curves) (based on mean GCSE grade).

Since we are looking at the variability of the proportion of C-A* students in centres, it would be expected that centres with abilities generating high proportions of candidates near the C/D borderline (e.g. centres with relatively weak cohorts) are more likely to experience higher levels of variability than very high or very low performing centres as they would be more sensitive to changes in their ability profiles. To see this, Table 7 lists the mean and standard deviation of the differences in centre proportions of C or above between 2016 and 2015 for three groups of centres based on their performance in 2015: the top 10%, the bottom 10% and the 20% of centres with the highest proportions of C/D candidates. As can be seen, the top performing centres had the lowest variability in outcomes compared to the other two groups (apart from science and additional science where the mean GCSE subject grade for the top performing centres was around the B/C boundary).

Table 7. Average of changes in centre proportions of C and above and the standard deviation for high and low performing centres.

Subject	Top 10 %			Middle 20 %			Bottom 10 %		
	Mean GCSE (2015)	Mean centre variability	SD	Mean GCSE (2015)	Mean centre variability	SD	Mean GCSE (2015)	Mean centre variability	SD
Maths	6.4	0.1	2.4	5.0	0.8	7.0	3.7	1.4	12.2
Add. Sci.	6.0	-0.3	5.8	5.0	-4.0	9.1	3.9	1.1	15.3
Science	5.6	-4.5	8.6	4.5	-3.3	12.5	3.3	-2.3	14.8
Biology	7.1	0.0	1.0	6.3	-0.3	4.2	5.3	-0.6	13.6
Chemistry	7.1	-0.1	1.9	6.3	-0.3	4.3	5.3	-3.4	15.8
Physics	7.1	0.0	1.1	6.3	-0.5	3.8	5.3	-4.2	14.5
English	6.7	-0.2	2.3	5.3	-2.9	6.2	4.1	-2.3	13.4
Eng. Lit.	6.7	-0.4	2.0	5.3	-2.3	8.1	4.2	3.8	11.9
Geography	6.7	0.2	3.8	5.4	-0.7	8.4	4.3	-2.3	15.5
History	6.8	0.0	3.3	5.4	-3.0	8.3	4.3	-1.8	13.2
French	7.0	-0.2	5.1	5.8	-0.7	12.0	4.7	-1.5	15.7
German	7.0	-0.4	3.7	5.9	-0.9	10.2	5.0	-9.0	19.4
Spanish	7.0	-0.3	4.6	5.8	-1.1	14.2	4.7	-4.3	17.8

Analysis based on prior attainment at KS2

Prior attainment at KS2 was also used as a measure of candidate's ability in the model (all the other variables remain unchanged), and the results are contained in Figures A12-A17 and Table A4 in the appendix. Overall, the model in this case fitted the data not as well as when mean GCSE grade was used. This is expected as KS2 tests were taken five years before GCSE and was not as good as mean GCSE grade in predicting candidate's performance at GCSE. However, mean KS2 fine level still has the largest influence on candidate's performance although the relative contributions from other factors have increased in comparison with the use of mean GCSE grade. The proportion of variance in the observed variability of centre GCSE outcomes that could be accounted for by the modelled variance ranges from 30% for history to 77% for English literature (see Table A4 in the appendix).

Conclusion

In this analysis, we have looked at the influence of candidates' background characteristics on their performance in examinations and the variability of outcomes in centres in successive years. We examined the effects using two approaches: analysis based on the average candidate profile of the centre and analysis based on the profiles of individual candidates within the centre.

Similar to previous work undertaken by Ofqual and others, we found that measures of socio-economic status have little or no bearing on centre variability. This indicates that the comparable outcomes approach to awarding does not have a systematic negative impact on centres with higher proportions of low socio-economic status candidates (similar to Ofqual, 2016).. Attainment at both GCSE and KS2 is found to be an important predictor of individual candidates' and individual centres' outcomes in any given year, although attainment at GCSE is a better predictor. How centre ability profile affects centre variability is not entirely clear, although centres with very high or low ability profiles are more likely to experience lower level of variability in outcomes than centres with ability profiles that produce large proportions of candidates with C or D grades. However, stable centres are not always those with low or high ability profiles. Centres with the most variability are those with a change in the ability of successive years of candidates.

Results from centre based analysis suggest that centres with a change in the number of students between years are more likely to experience variability in outcomes. Also, whether a centre was stable or variable in one year is a predictor of stability or variability in the following year. This means that centres who are stable in one year are likely to be stable the following year. However, centres who experienced positive volatility in one year are likely to experience negative volatility in the next. This is probably because such centres have a high proportion of candidates who are clustered around the C/D borderline (rather than, for example, clustering around grade B) and so more year-on-year variability is inevitable, even if there is a small change in the entry size and/or ability profile. A comparison between centre based models suggests that changes to the average profile of students between 2015 and 2016 is more significant than the actual 2015 and 2016 profile of candidates. Whilst there are some subject differences, only the change in attainment at GCSE or KS2, the previous stability classification of the centre and the percentage change in cohort size are significant factors. When using the dynamic model, around 70% to 75% of centres are correctly classified as stable or negative unstable. Based on this model, a significant amount of variability is not predictable. It is likely that there are some features within centres that are difficult to 'capture' and model, such as changes in timetabling or to staff etc.

Results from candidate based analysis indicate that the large majority of the observed variability in centre level outcomes can be attributed to changes in the ability profiles and the other characteristics of successive years of candidates within the centres. However, change in ability profile has the largest influence on centre variability. Only a small proportion of the observed centre variability is found to be associated with the indeterminacy of the examinations system which will always be likely to exist.

References

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 13 : Receiver operating characteristic curves, *8*(6), 508–512.

Bramley, T. and Benton, T. (2015) Volatility in exam results. Cambridge Assessment Research Report, Cambridge, UK: Cambridge Assessment

Crawford, C. and Benton, T (2017) Volatility happens: Understanding variation in schools' GCSE results. Cambridge Assessment Research Report, Cambridge, UK: Cambridge Assessment. Accessed at: http://www.cambridgeassessment.org.uk/Images/372751-volatility-happens-understanding-variation-in-schools-gcse-results.pdf

Ofqual. (2015). Variability in GCSE Results for Individual Schools and Colleges 2012 to 2015.

Ofqual (2016). What causes variability in school-level GCSE results year-on-year? Coventry, UK. Accessed at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/518409/Variability_in_Individual_Schools_and_Colleges_2016.docx_-_FINAL.pdf

Ofqual (2017). Interactive Graphs: variability in results for schools and colleges in 2017: GCSE. Accessed at: https://analytics.ofqual.gov.uk/apps/2017/GCSE/CentreVariability/

Ofqual (2018). Interactive Graphs: variability in results for schools and colleges in 2018: GCSE. Accessed at: https://analytics.ofqual.gov.uk/apps/2018/GCSE/CentreVariability/

Appendix

Table A1.Correlations of candidate's subject grade with background characteristics for the 13 subjects from the 2016 dataset.

Subject	Mean GCSE	Mean KS2	First Lang.	SEN	IDACI	FSM	Gender	Centre type
Maths.	0.93	0.60	0.02	-0.24	-0.18	-0.16	-0.08	0.06
Add. Sci.	0.90	0.53	0.03	-0.08	-0.18	-0.11	-0.09	0.12
Science	0.81	0.56	0.03	-0.13	-0.18	-0.12	-0.14	0.13
Biology	0.90	0.52	0.06	-0.09	-0.16	-0.10	-0.08	0.12
Chemistry	0.86	0.67	-0.05	-0.30	-0.22	-0.18	-0.21	0.12
Physics	0.86	0.63	-0.02	-0.28	-0.21	-0.17	-0.22	0.11
English	0.92	0.67	-0.03	-0.25	-0.27	-0.19	-0.12	0.12
Eng. Lit.	0.82	0.56	0.06	-0.10	-0.12	-0.10	-0.15	0.13
Geog.	0.91	0.63	0.01	-0.21	-0.23	-0.18	-0.11	0.11
History	0.90	0.74	0.01	-0.32	-0.22	-0.19	-0.01	0.12
French	0.89	0.54	0.04	-0.07	-0.17	-0.11	0.00	0.12
German	0.93	0.62	-0.01	-0.29	-0.19	-0.17	-0.07	0.05
Spanish	0.79	0.53	0.07	-0.13	-0.16	-0.11	-0.15	0.12

Table A2. Centre stability associated with each subject. The mean GCSE is calculated by converting grades A* to G into a numeric format (where A* = 8, A = 7, …, G = 1) and taking the mean value.

Subject	Stability	Number of centres	Mean 2016 GCSE grade	Mean GCSE grade change
Mathematics	Negative	334	4.7	-0.4
Mathematics	Stable	1,824	5.3	0.0
Mathematics	Positive	340	5.1	0.5
Additional Science	Negative	291	4.3	-0.9
Additional Science	Stable	1,466	4.8	-0.1
Additional Science	Positive	275	4.8	0.5
Science	Negative	140	4.0	-0.8
Science	Stable	698	4.4	-0.1
Science	Positive	135	4.9	0.6
Biology	Negative	84	5.6	-0.5
Biology	Stable	717	6.4	0.0
Biology	Positive	99	6.1	0.5
Chemistry	Negative	93	5.5	-0.5
Chemistry	Stable	758	6.4	0.0
Chemistry	Positive	89	6.0	0.5
Physics	Negative	93	5.4	-0.6
Physics	Stable	782	6.4	0.0
Physics	Positive	93	6.1	0.5
English	Negative	159	4.9	-0.5
English	Stable	1,040	5.4	0.0
English	Positive	163	5.4	0.5
English literature	Negative	177	4.9	-0.7
English literature	Stable	1,224	5.6	-0.1
English literature	Positive	165	5.4	0.6
Geography	Negative	223	4.6	-0.8
Geography	Stable	1,182	5.4	-0.1
Geography	Positive	205	5.3	0.7
History	Negative	169	4.5	-0.9
History	Stable	862	5.3	-0.1
History	Positive	174	5.3	0.6
French	Negative	151	4.7	-0.7
French	Stable	739	5.4	0.0
French	Positive	139	5.4	0.7
German	Negative	51	4.8	-0.7
German	Stable	261	5.5	0.0
German	Positive	46	5.6	0.6
Spanish	Negative	77	4.7	-0.7
Spanish	Stable	395	5.6	0.0
Spanish	Positive	75	5.5	0.7

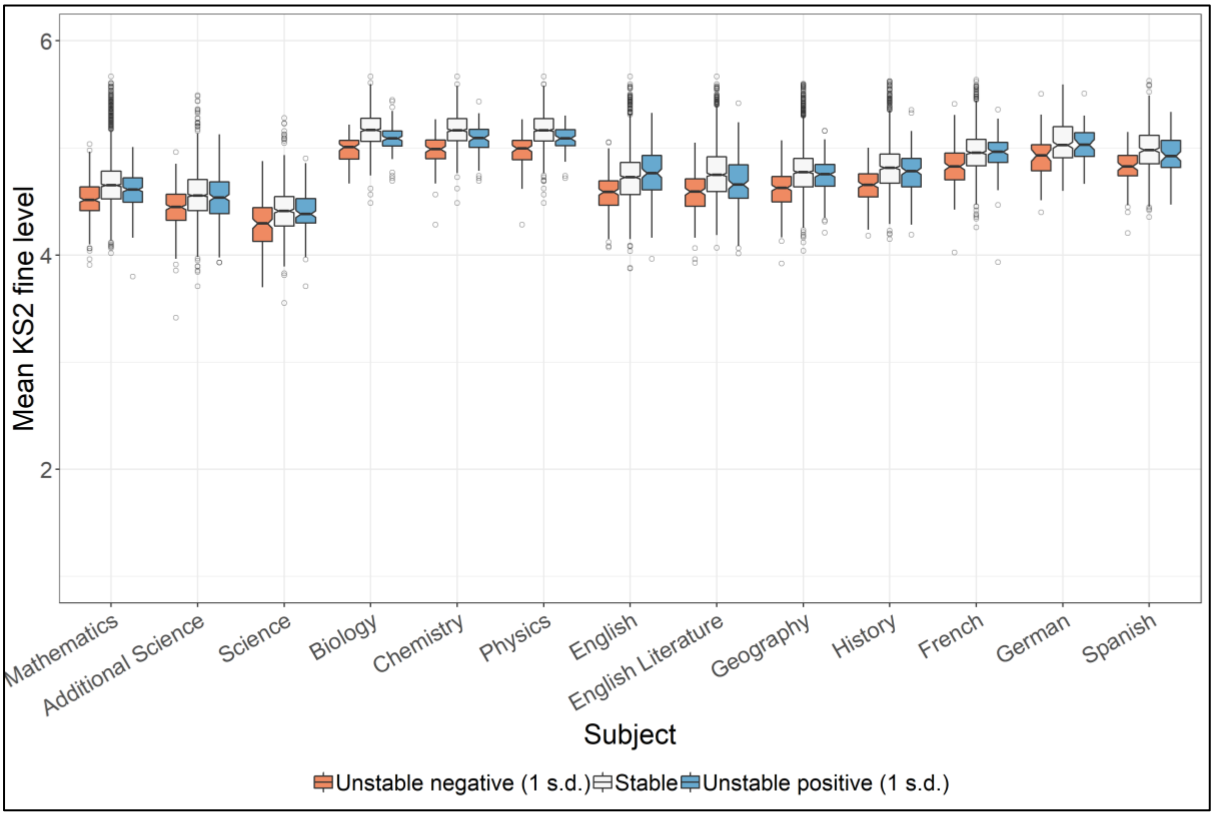

Figure A1. Distributions of centre average KS2 fine level for the 2016 cohort.

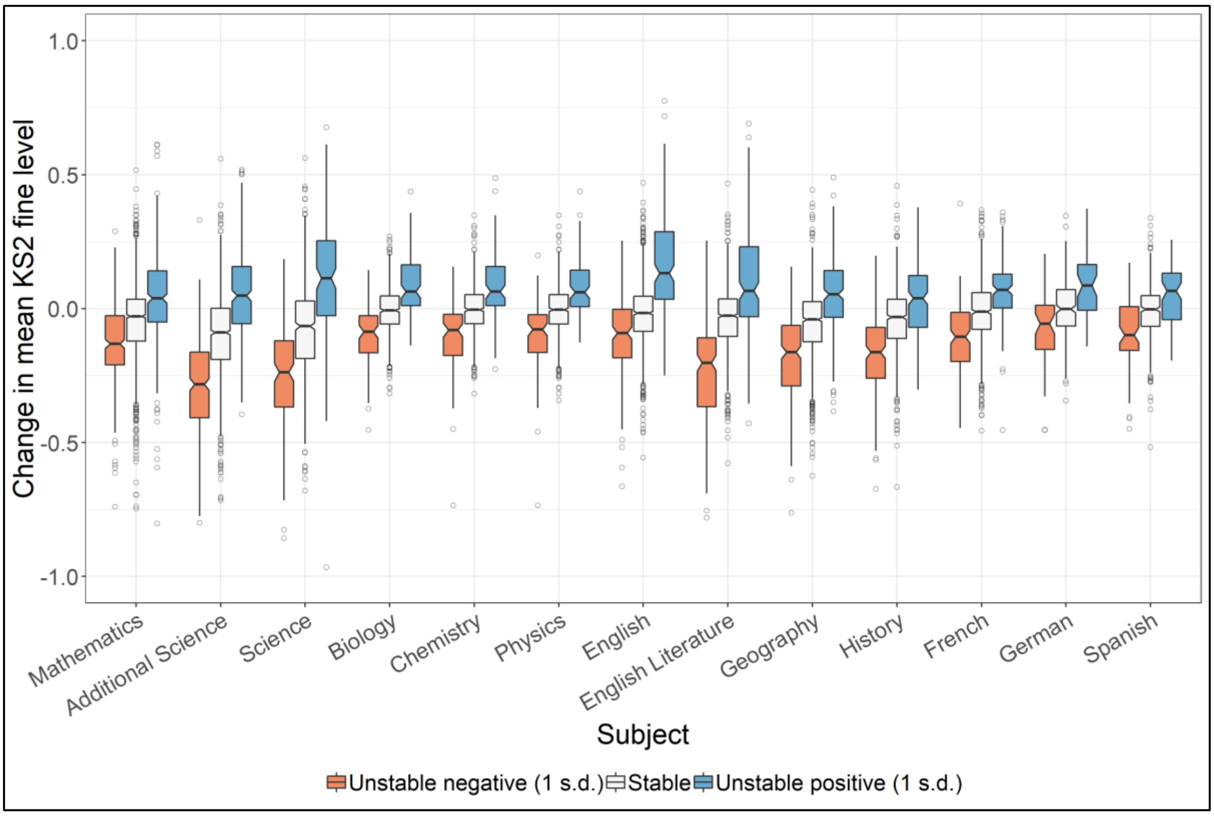

Figure A2. Distributions of changes in centre average KS2 fine level between the 2015 and 2016 cohorts.

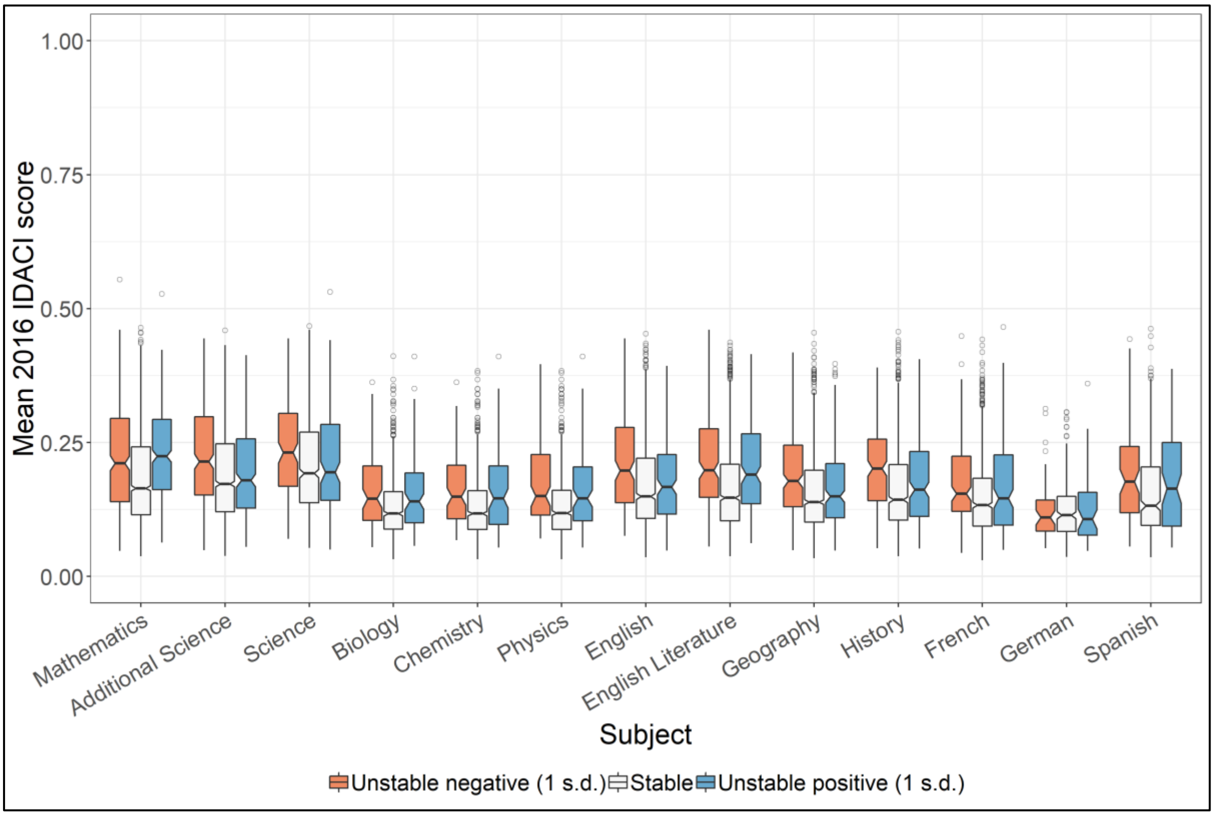

Figure A3. The mean IDACI score of the 2016 cohort.

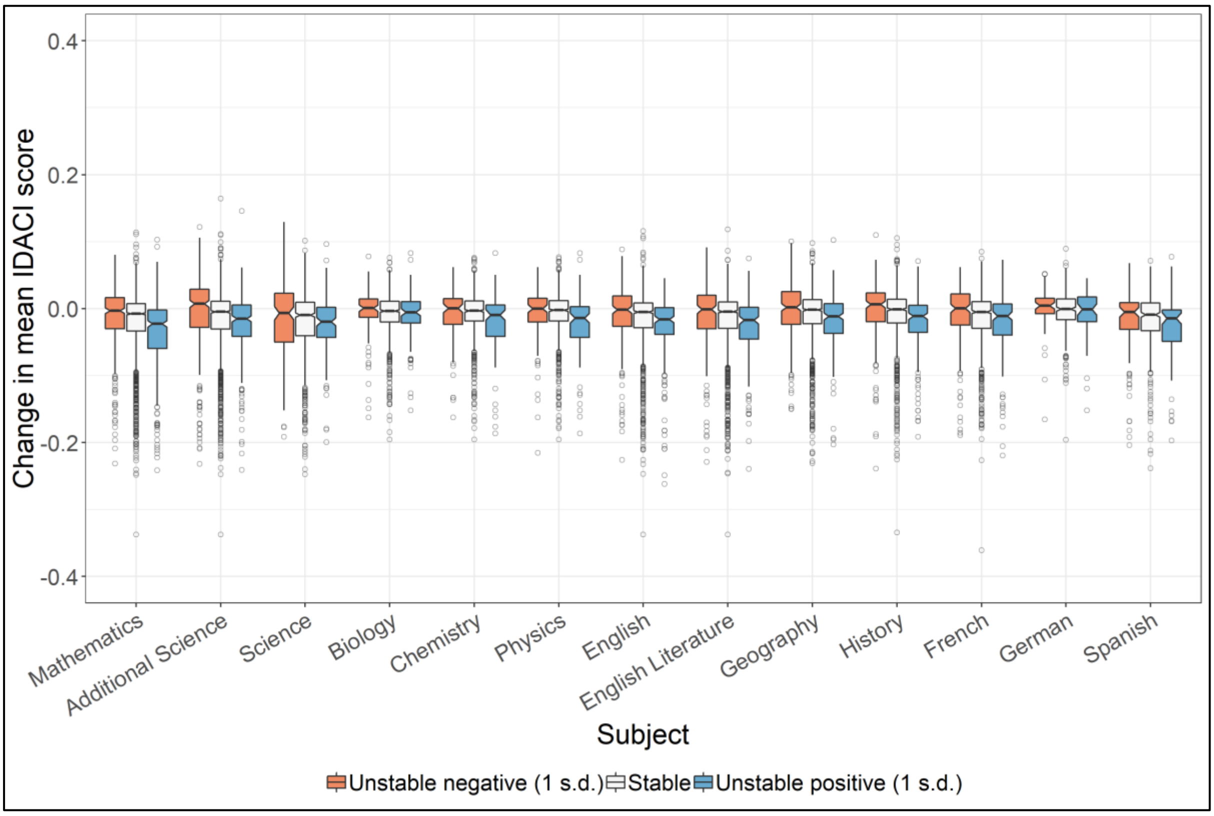

Figure A4. The change in the mean IDACI

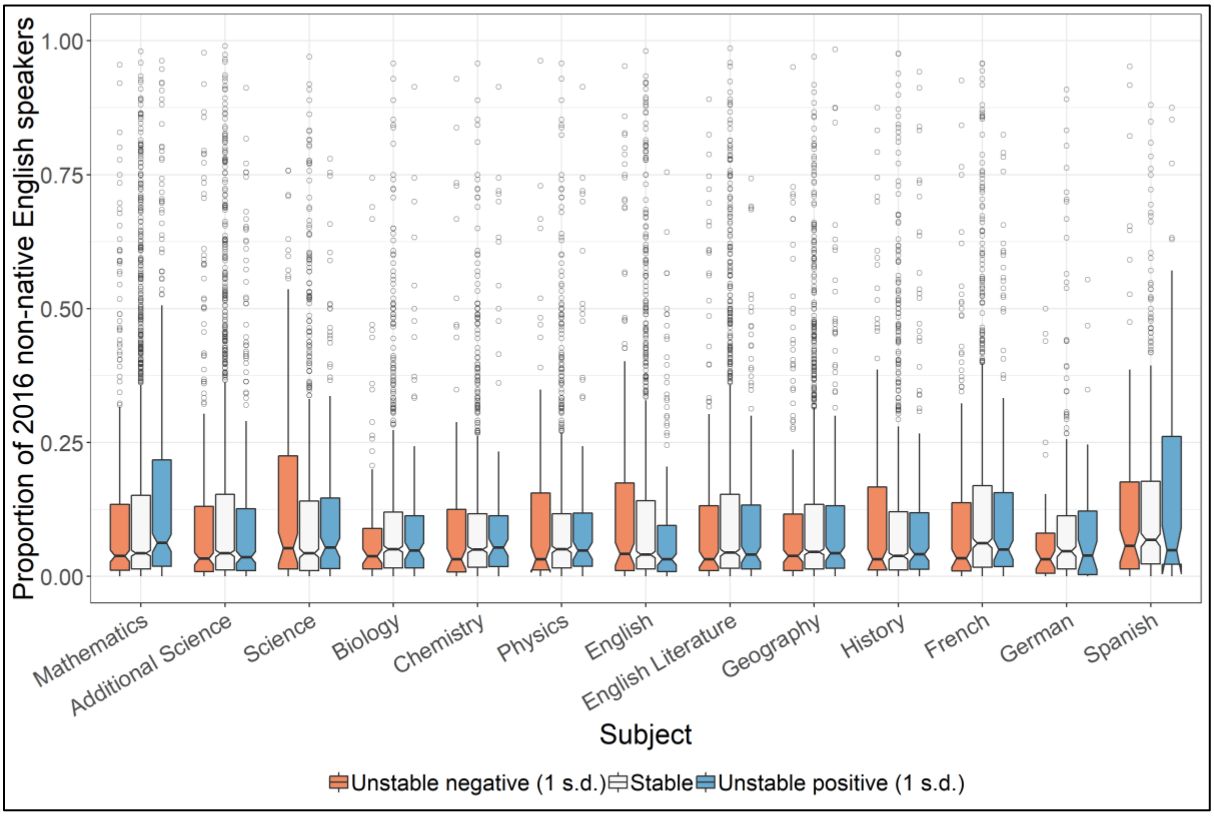

Figure A5. Proportion of non-native English speakers in the 2016 cohort.

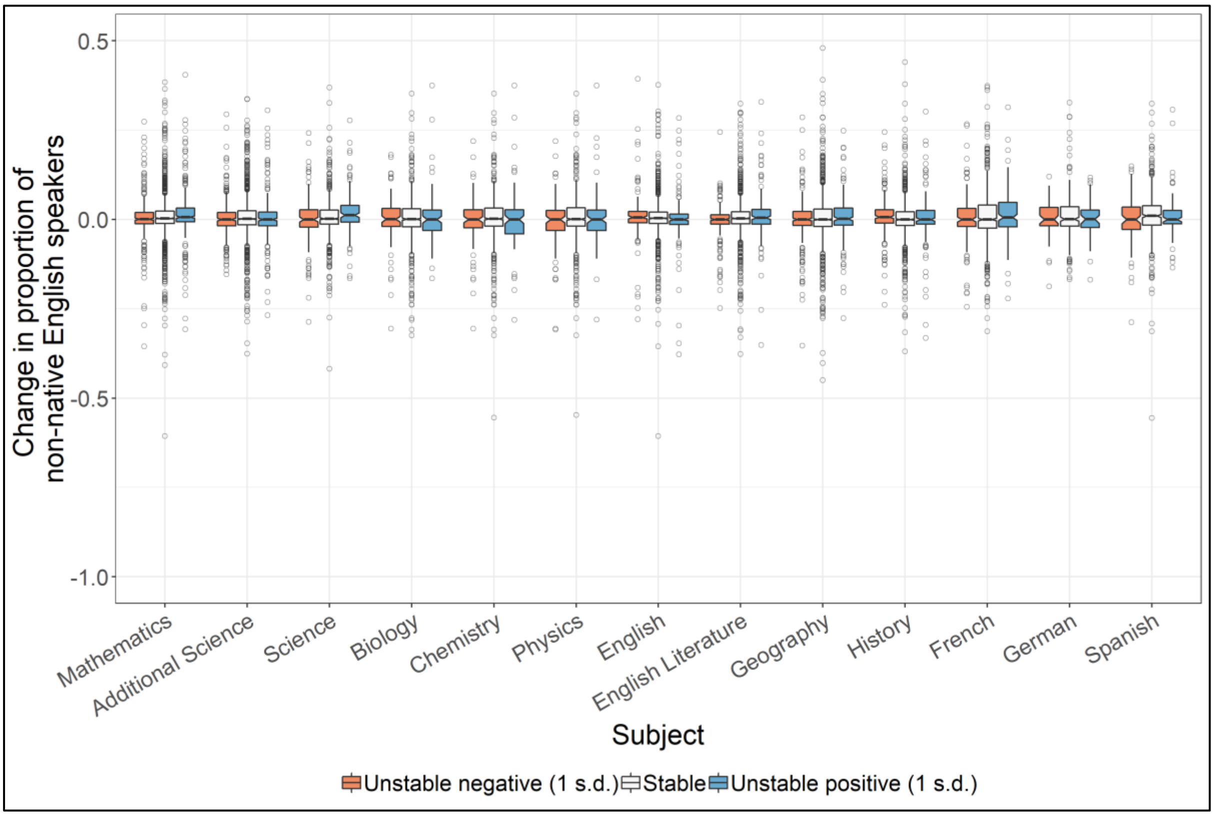

Figure A6. Change in the proportion of non-native English speakers.

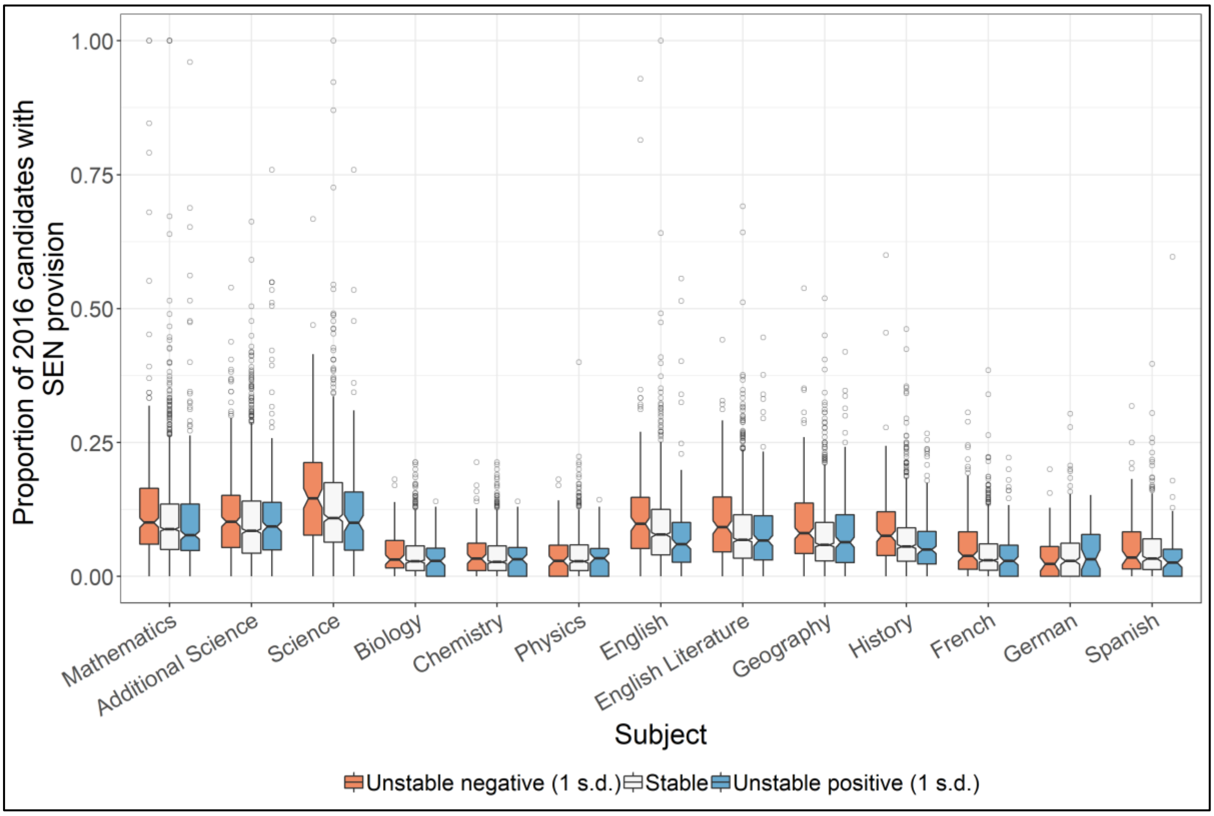

Figure A7. Proportion of candidates with SEN provision in 2016 cohort.

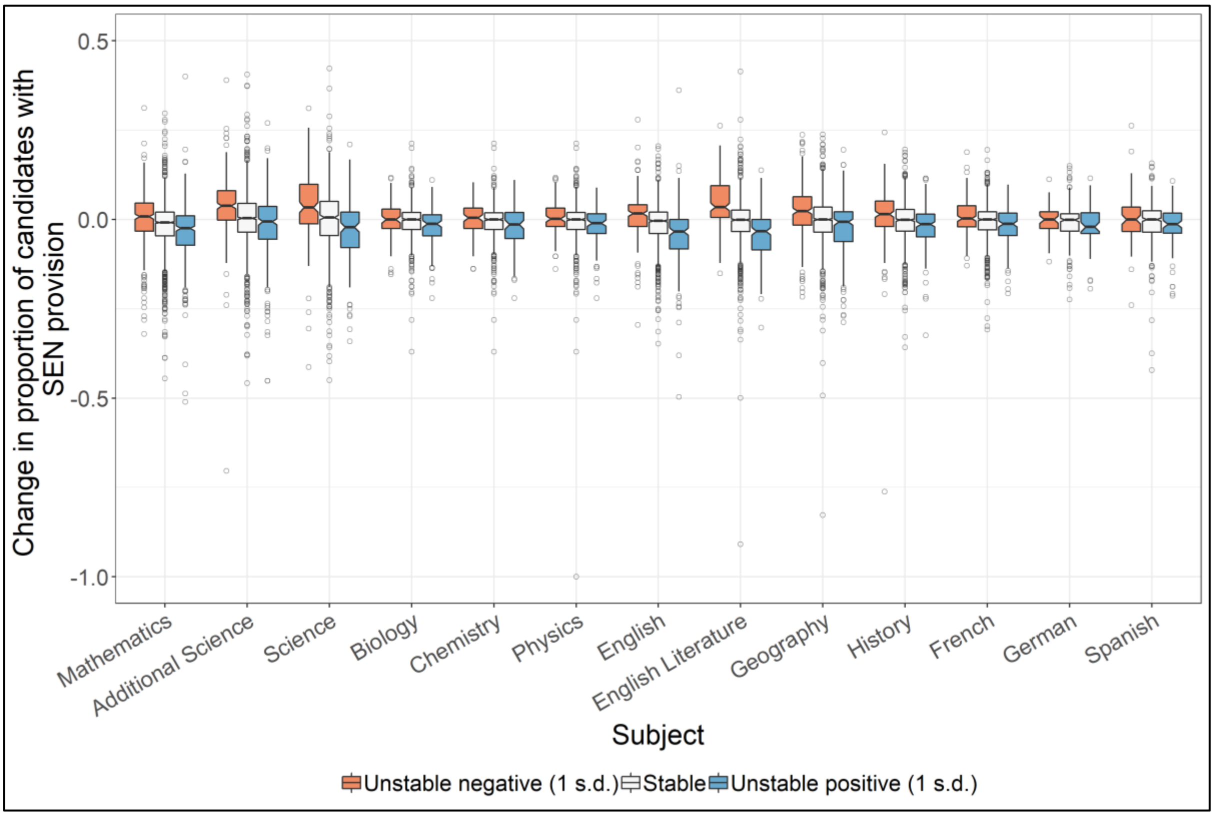

Figure A8. Change in proportion of candidates with SEN provision.

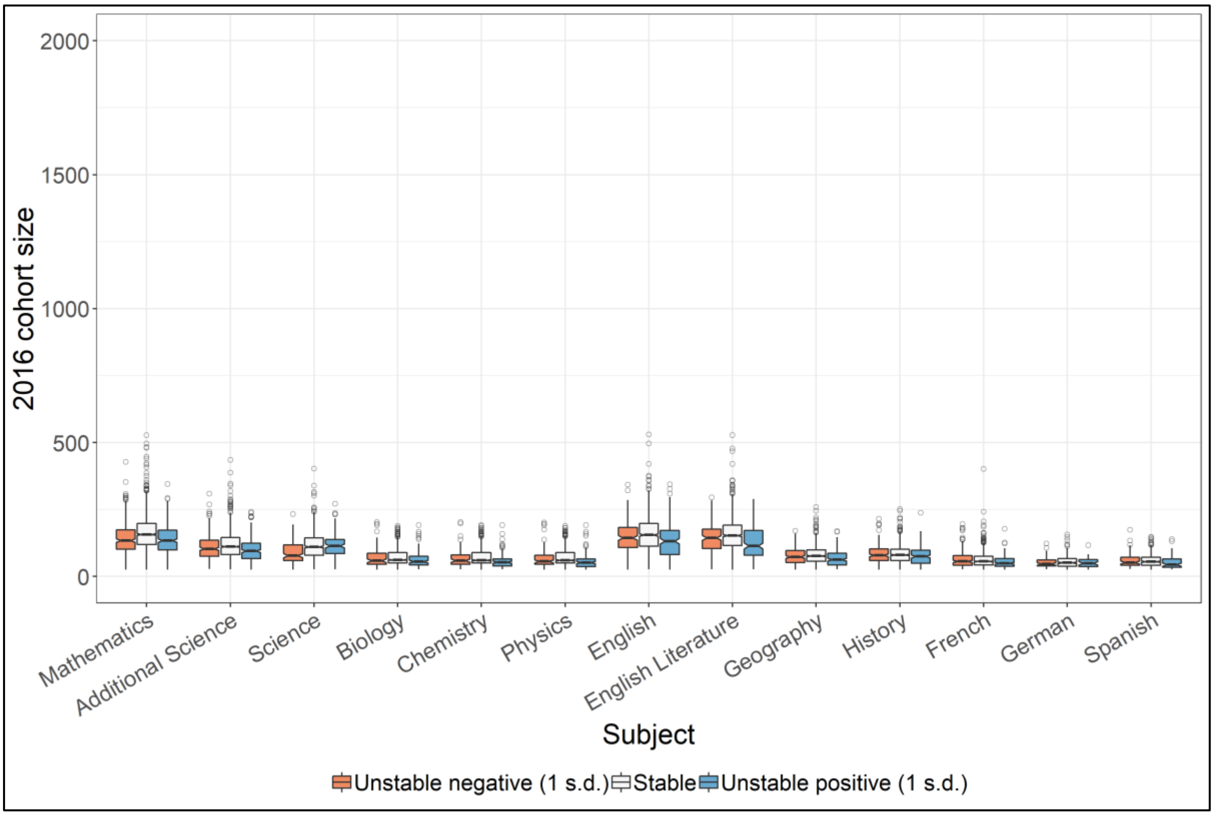

Figure A9. Mean cohort size of the 2016 cohort.

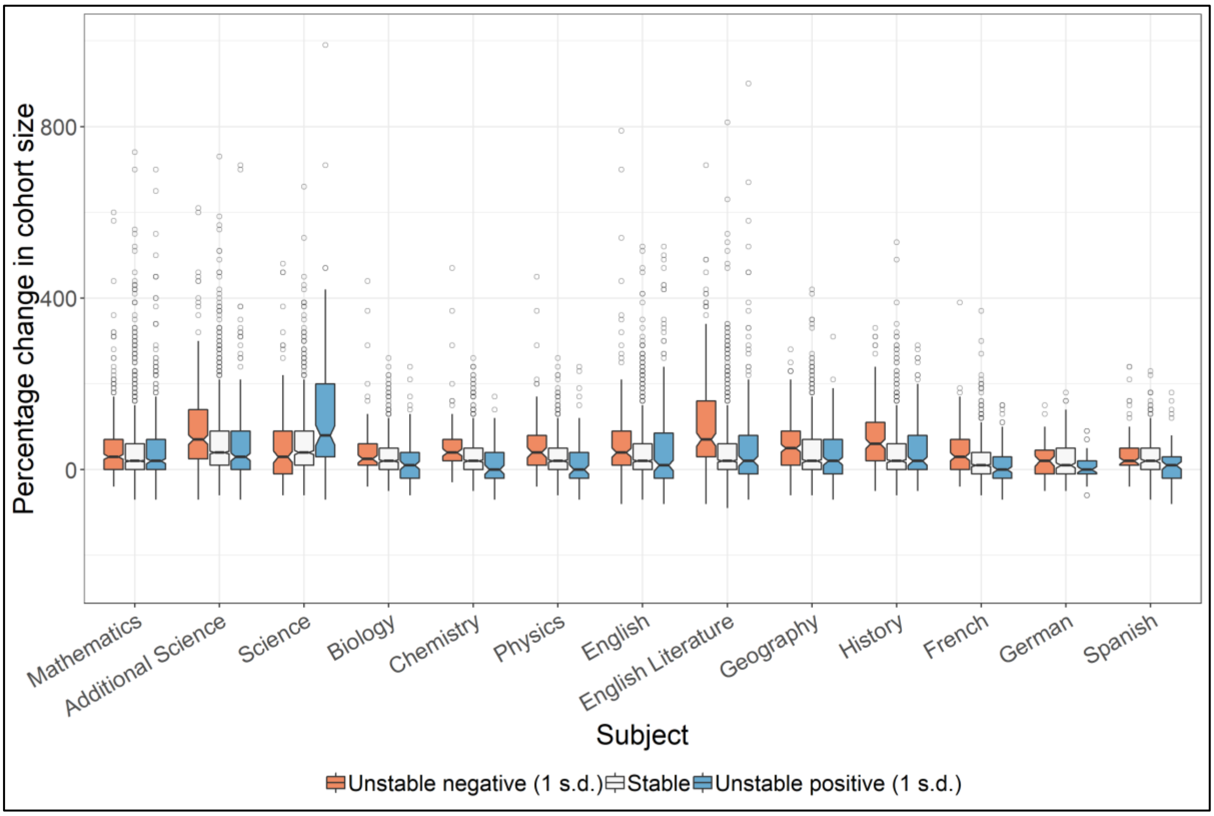

Figure A10. Percentage change in the cohort size.

Table A3. Significance of each factor included in equation (2), broken down by subject (based on average centre mean KS2 fine level).

Subject	Intercept	Centre type	FSM CHG	Male CHG	IDACI CHG	Ability CHG	Size CHG	Neg. 2015	Pos. 2015	SEN CHG	EAL CHG
Maths.	*		*			*		*	*		
Add. Sci.	*		*			*	*		*	*	
Science	*					*		*	*		
Biology	*					*			*		
Chemistry	*					*			*		
Physics	*					*	*		*		
English	*			*		*	*			*	
Eng. Lit.	*			*		*			*	*	*
Geog.	*				*	*	*		*		*
History	*					*	*	*	*		*
French	*		*	*		*	*		*		
German	*			*		*			*		
Spanish	*					*			*		

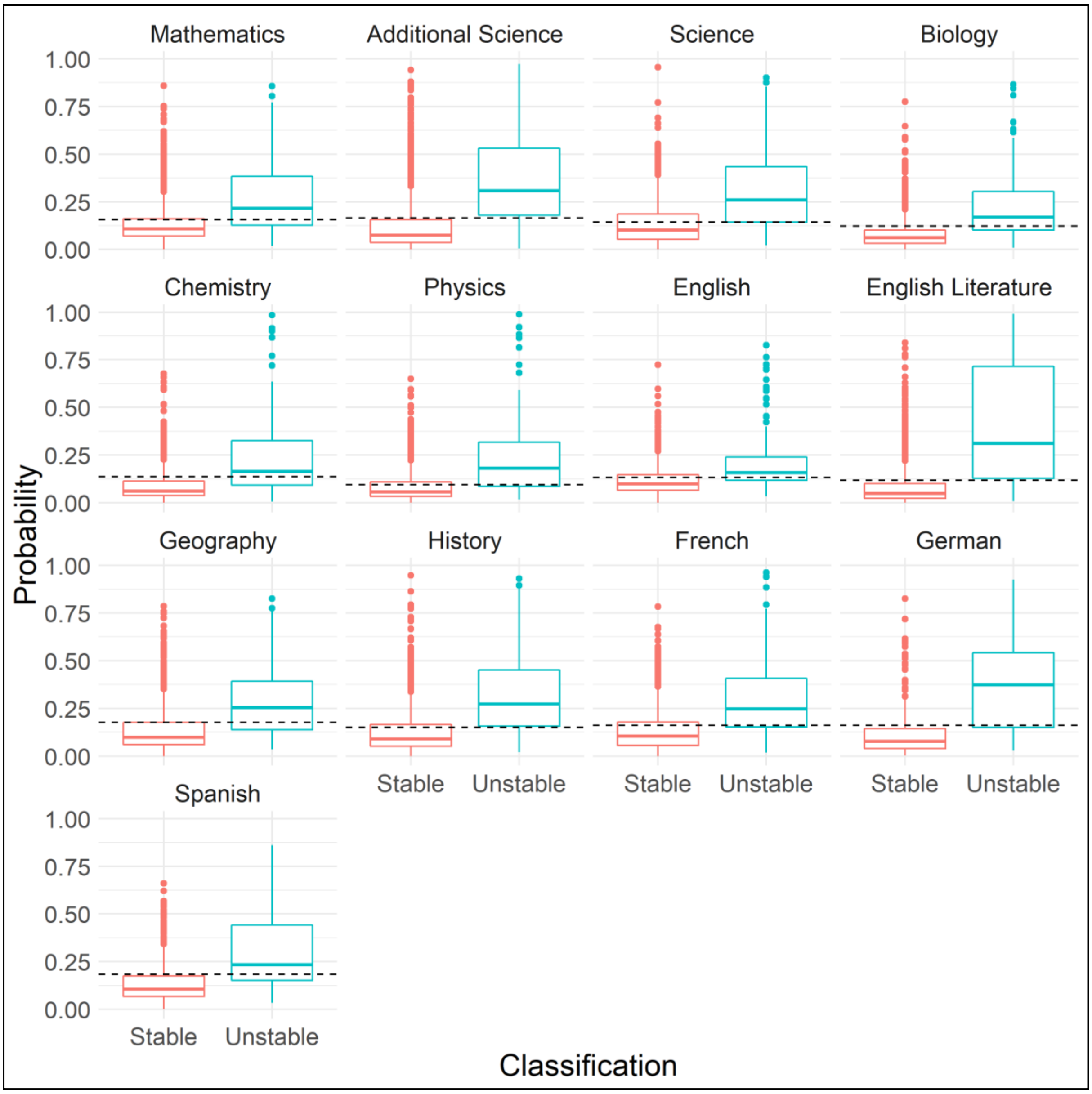

Figure A11. Probability of stable and negative unstable centres being classified negative unstable (based on centre average mean KS2 fine level). The classification threshold is given by the black line.

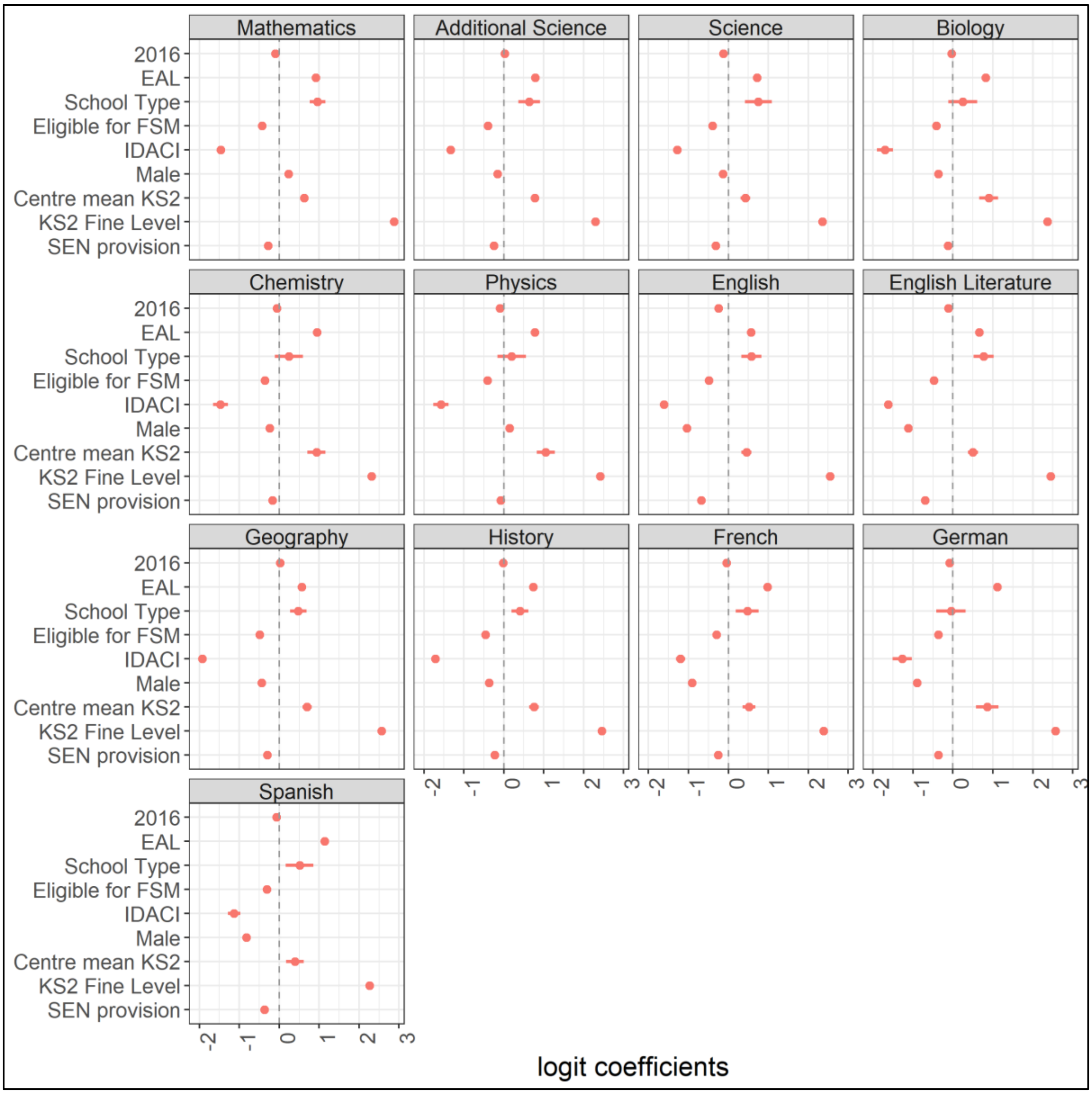

Figure A12. Distributions of the estimated model coefficients and the associated confidence intervals for the two-level logistic regression model using mean KS2 fine level as the measure of candidate's ability for the 13 subjects studied.

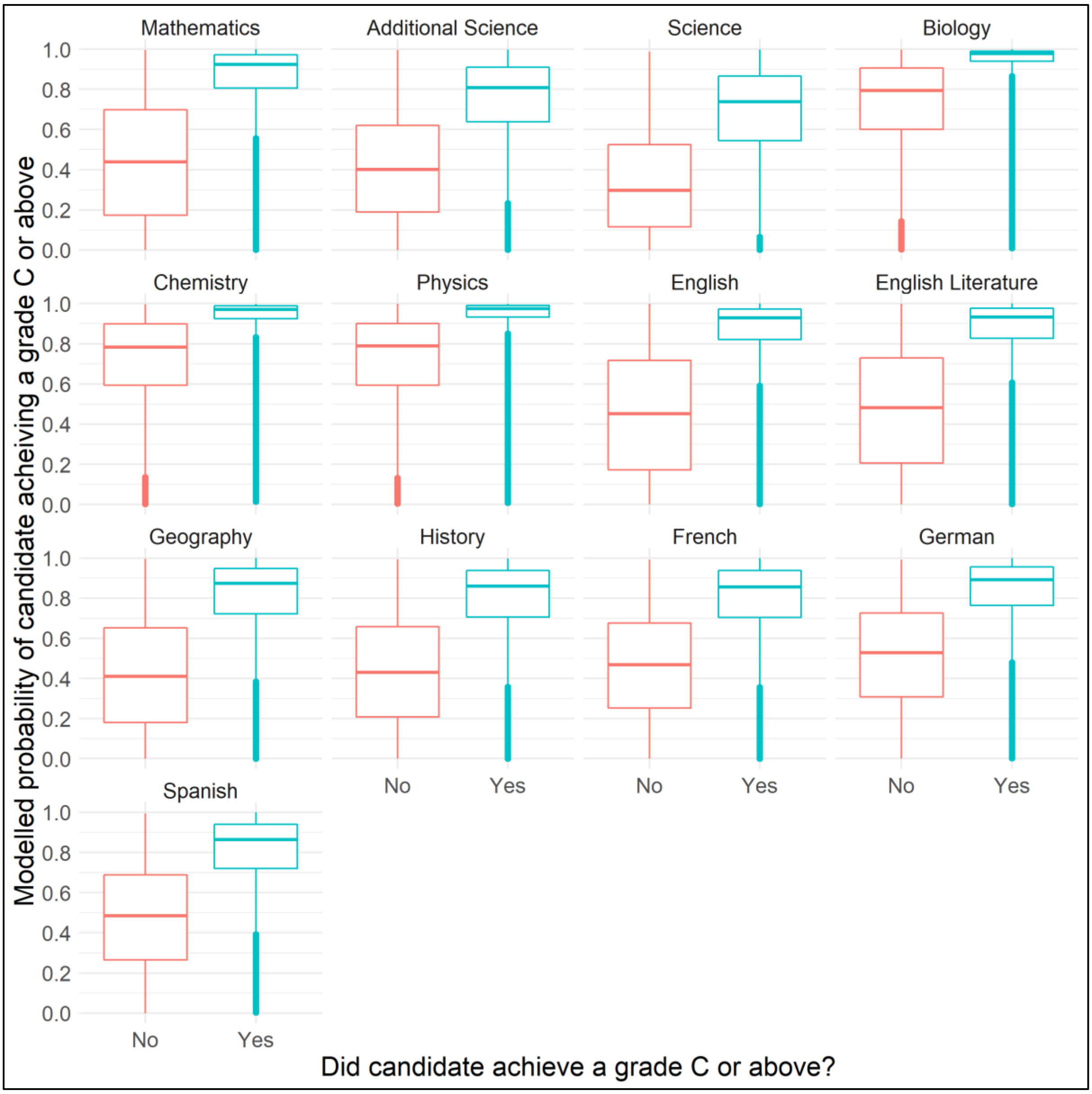

Figure A13. Distributions of model predicted probability for candidates receiving a grade C or above and the probability for those receiving a D or below based on their background characteristics for the 13 subjects studied (based on mean KS2 fine level).

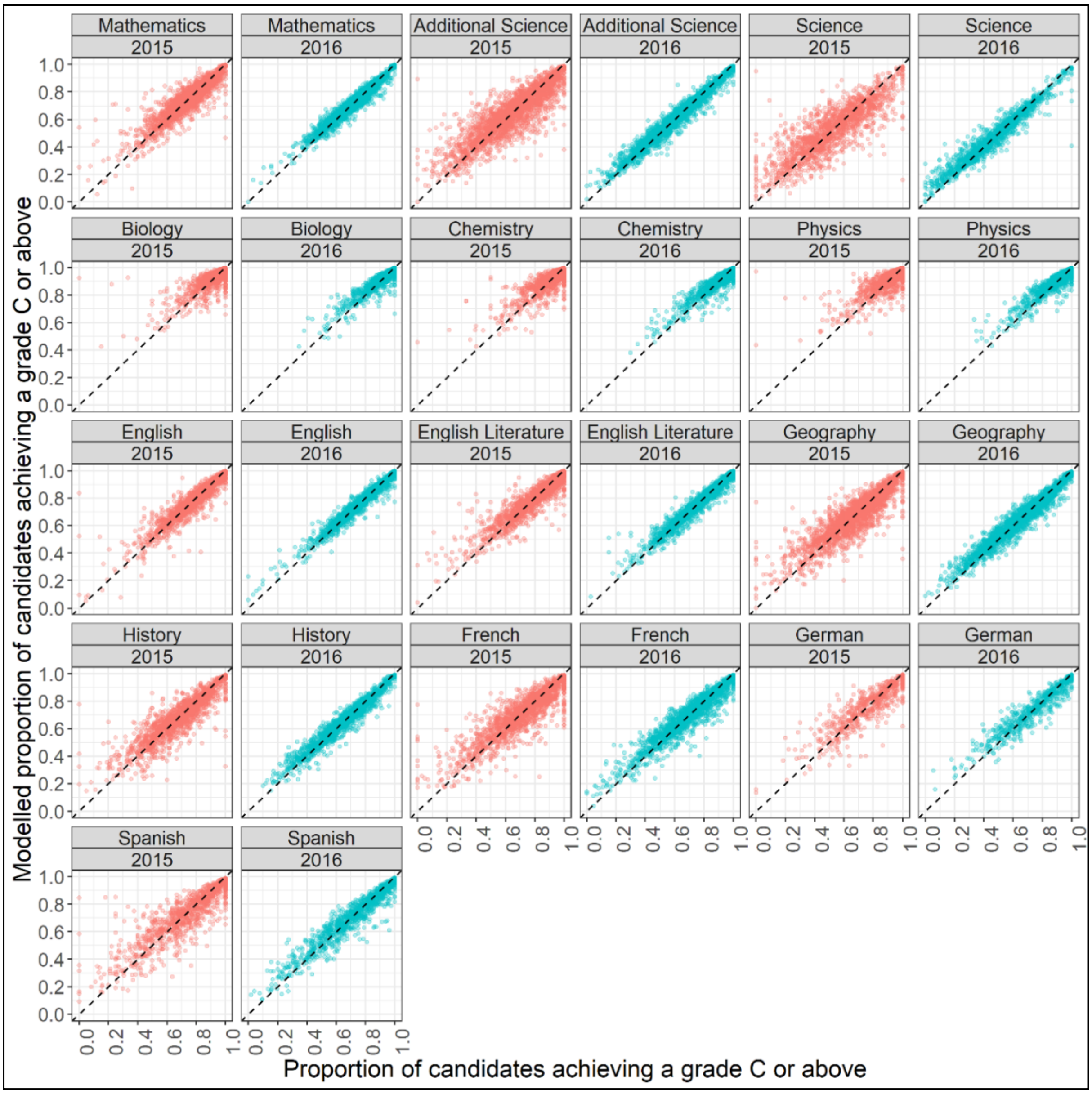

Figure A14. Comparison of modelled centre level proportion of candidates receiving a grade C or above and observed proportion in 2015 and 2016 for the 13 subjects (based on mean KS2 fine level).

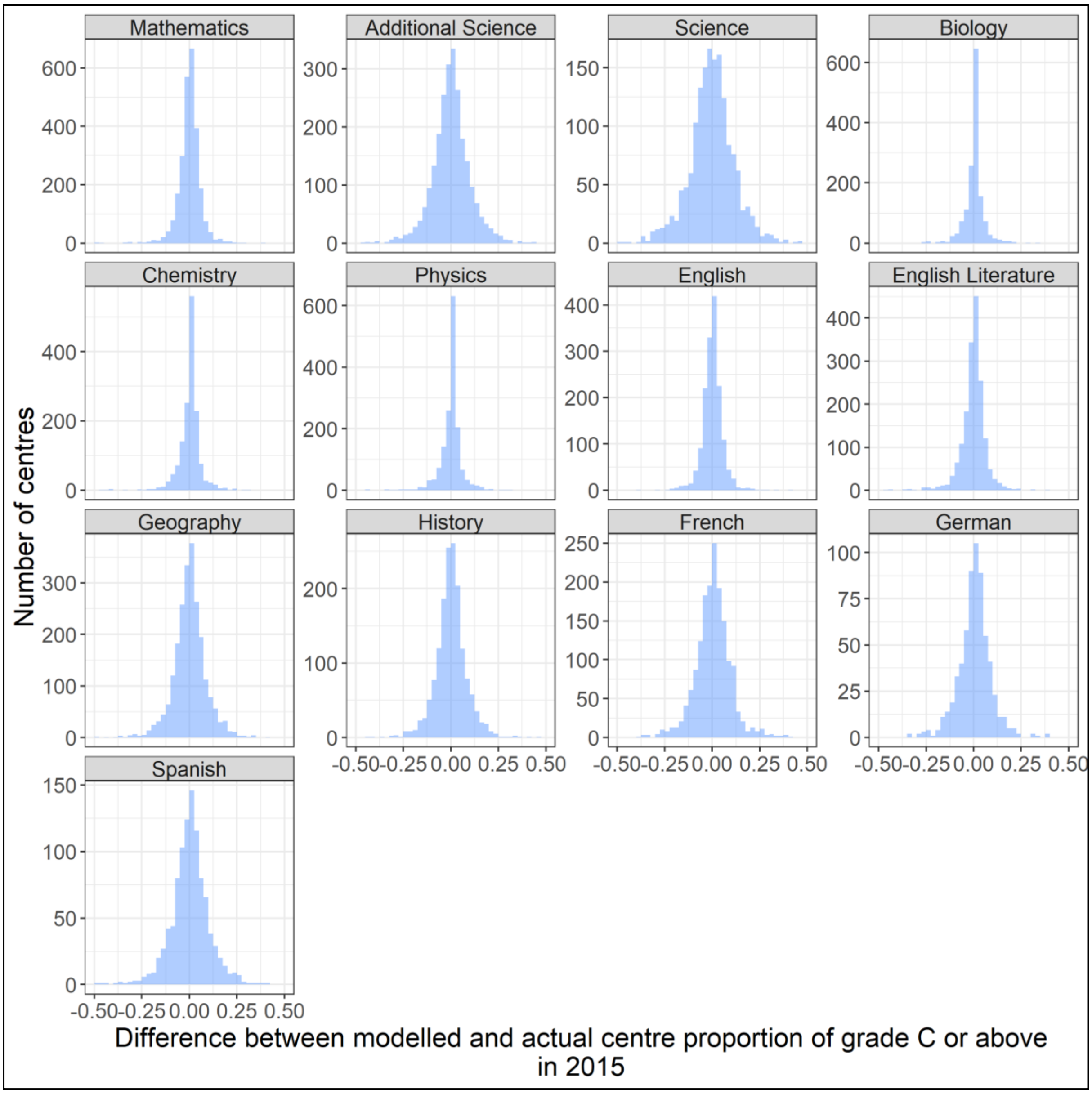

Figure A15. Frequency distribution of the difference between observed centre level proportion of candidates achieving a C or above and the model predicted proportion for the 13 subjects in 2015 (based on mean KS2 level).

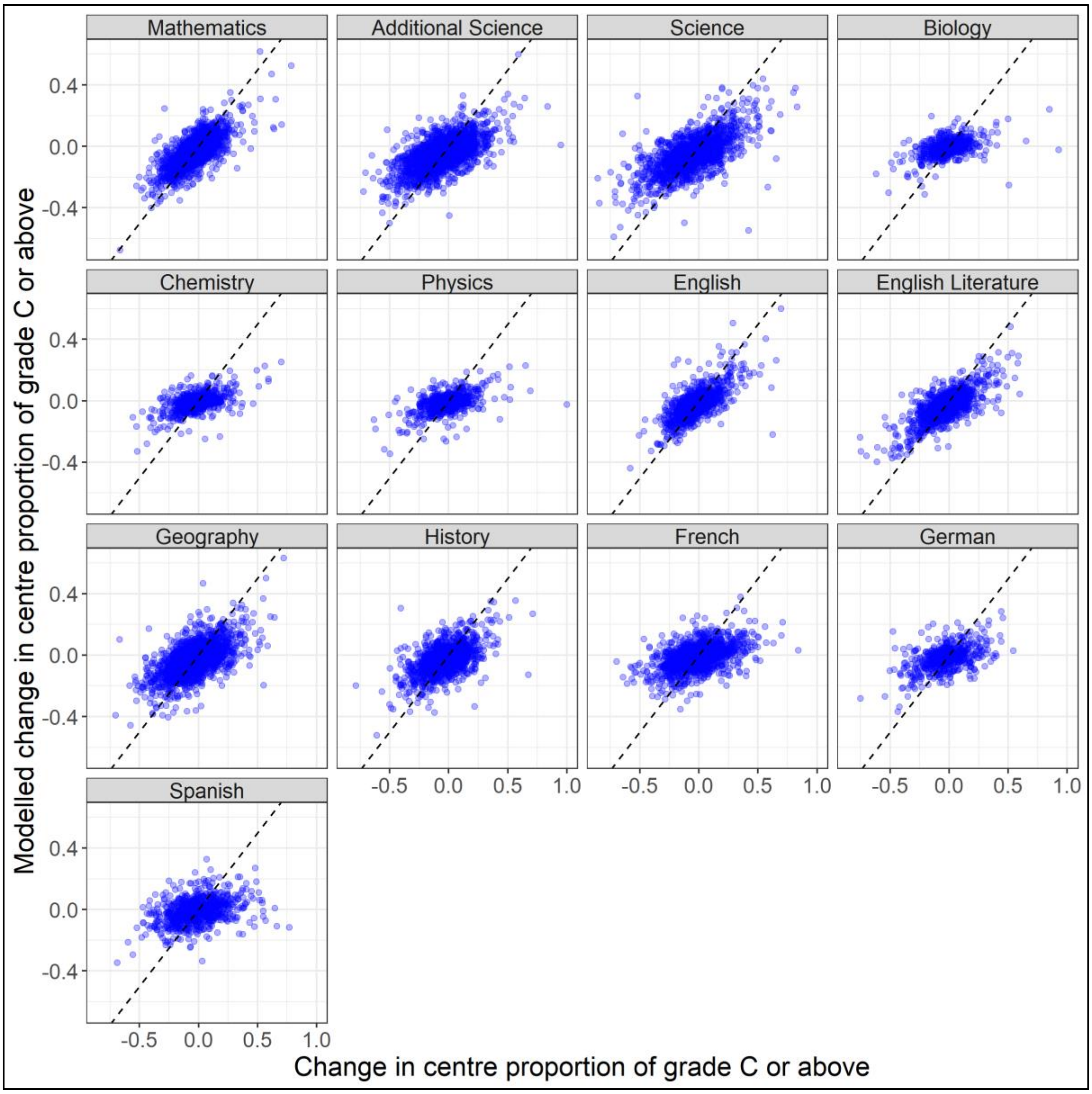

Figure A16. Comparison of observed centre level variability in outcomes between 2016 and 2015 and the model predicted variability (based on mean KS2 level).

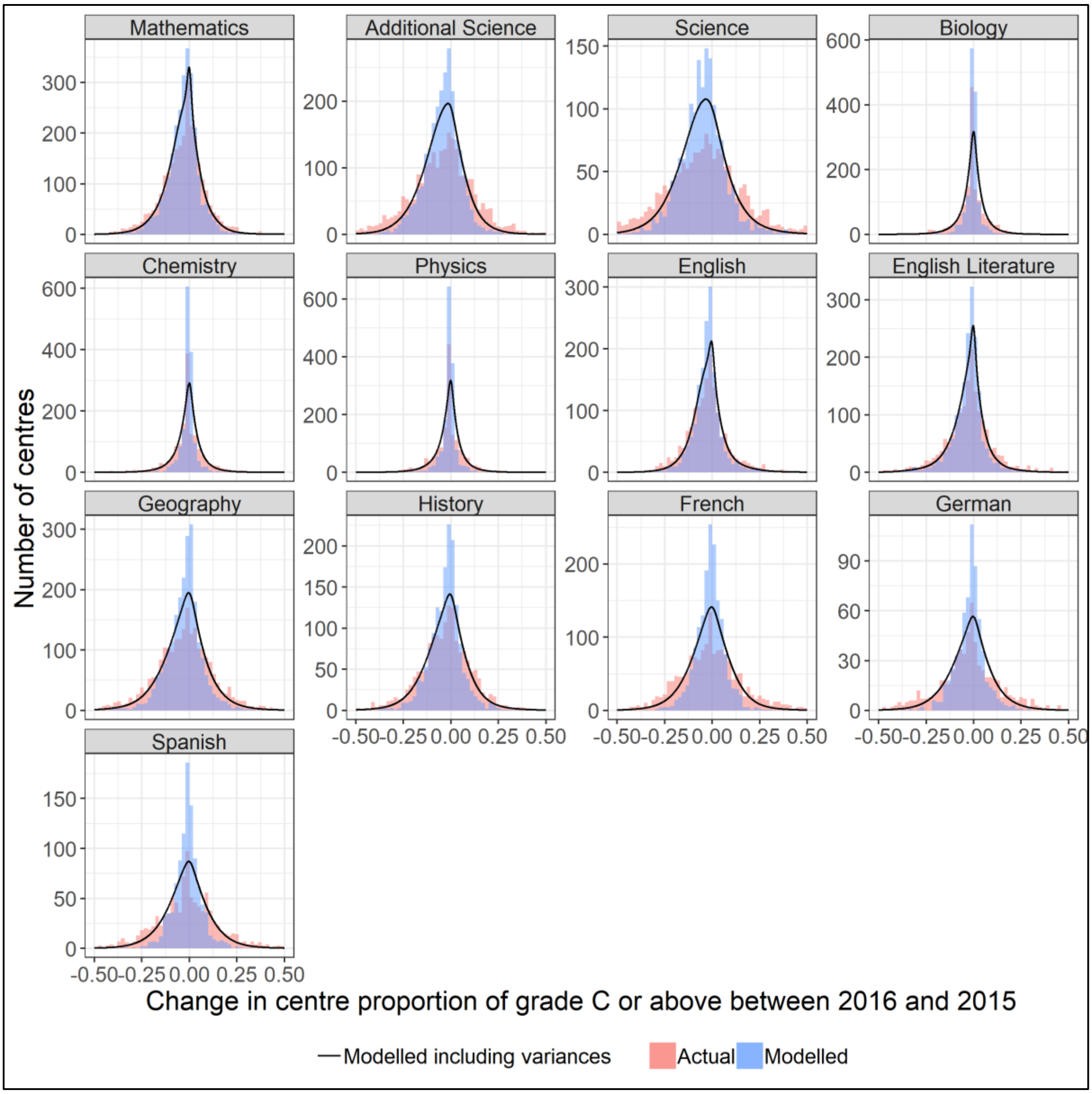

Figure A17. Comparison of the frequency distributions of the observed centre level changes in proportions of candidates achieving C or above between 2016 and 2015 (red bars) and the model predicted changes (blue bars) and the modelled probability distributions against the level of changes for all centres (the smoothed black curves) (based on mean KS2 level).

Table A4. Correlations between model predicted changes in centre GCSE outcomes and observed changes and between observed changes and changes in centre average mean KS2 fine level.

Subject	Modelled changes in centre outcomes vs observed changes	Observed changes centre outcomes vs changes in average mean KS2 level
Maths.	0.85	0.33
Add. sci.	0.71	0.06
Biology	0.72	0.12
Chemistry	0.82	0.40
Physics	0.88	0.25
Science	0.73	0.09
English	0.74	0.29
Eng. lit.	0.68	0.15
Geog.	0.55	0.30
History	0.72	0.61
French	0.71	0.13
German	0.85	0.32
Spanish	0.69	0.08

Published by:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344
public.enquiries@ofqual.gov.uk
www.gov.uk/ofqual

October 2018 **Ofqual/18/6432**