

ANALYSIS AND DECISIONS DOCUMENT

Grading standards in GCSE French, German and Spanish

Evaluating the evidence for an adjustment to
grading standards in GCSE French, German and
Spanish on the basis of inter-subject comparability

Contents

Executive Summary	3
Introduction.....	9
Our evidence.....	9
A summary of the criteria.....	10
The evidence under Criterion A.....	12
The evidence under Criterion B.....	20
The evidence under Criterion C.....	29
The evidence under Criterion D.....	41
Decisions.....	49
Implementation.....	54
Impact Assessment.....	55
Equality Analysis.....	55
Regulatory Impact and Growth.....	56
References	57

Executive summary

We have investigated whether GCSEs in French, German and Spanish are severely graded in comparison to GCSEs in other subjects. We have concluded that grading standards in GCSE French and German, but not Spanish, should be adjusted. The evidence we have considered and the criteria we have used to come to our decision are set out in this document.

We gathered evidence from a range of sources, including:

- Statistical evidence (including measures of relative subject difficulty)
- National and international trends in modern foreign language entries
- Stakeholders' views
- Awarding organisations' views, including the views of their awarders
- Academic papers
- Benchmarking against the Common European Framework of Reference for Languages (CEFR)

We also considered the potential impact on society if GCSEs in these three languages were severely graded.

We considered the evidence for each subject against a set of criteria to determine whether there was a 'compelling case' to adjust grading standards. We used the set of criteria because we determined that no single source of evidence could definitively demonstrate that grading standards in these subjects were out of line with those of other GCSEs; the individual sources of evidence all have inherent limitations.

In this document we summarise our analyses of the evidence by each criterion, and set out our decisions and the action we will take as a result.

Criterion A – Statistical Evidence

Our first criterion is concerned with statistical evidence of potential grading severity. While ostensibly quite compelling, there are significant limitations with statistical measures of the relative difficulty of different subjects to the extent that we do not believe it can be relied on in isolation and must be treated with caution.

We considered the extent to which different measures suggested that these subjects were more 'difficult' in statistical terms than other GCSEs, and whether this had been consistently the case for a significant period of time. We also considered whether multiple sources of statistical evidence pointed in the same direction. We looked for consistency over time, as the apparent 'statistical difficulty' of a subject in statistical terms can vary year-on-year due to factors unconnected to the subject itself, such as changes in entry.

Statistical analyses, using a variety of different methods, suggest that these subjects are generally more 'difficult' overall than the average for all GCSEs. Repeating these

analyses on National Pupil Database data from 2002 – 2018 suggests that this relative severity has been present for a significant period of time. This was particularly the case for GCSE French and German. For Spanish there was a less consistent picture and evidence that, at some grades, the subject was, in fact, easier than the average of all GCSE subjects. French and German also consistently appear to be among the five hardest subjects at GCSE; Spanish is closer to the mean GCSE difficulty in some years. We therefore found a statistical case for adjusting standards in French and German, but not in Spanish.

Criterion B – The Impact of Severe Grading Standards

The second criterion concerned evidence of negative impacts potentially arising if GCSEs in French, German and Spanish were out of line with those of other subjects. We focused on trends in A level and GCSE entries, as well as indications of skills shortages and teacher supply. We sought to understand the extent to which such shortages arose from grading standards, rather than for other reasons.

GCSE and A level entries for French and German have declined significantly over the past two decades, but Spanish entries have been steadily increasing over the same period. Rates of progression from GCSE to A level have remained relatively stable in all three subjects, but are lower than those in other optional GCSE subjects such as history and geography. There was a small increase in GCSE entry in 2013 in each of the three languages, but French and German have since continued to decline – although the most recent figures suggest German may be stabilising.

We found evidence that the recruitment of teachers in these subjects is becoming more difficult (both in terms of Initial Teacher Trainee applications from the UK and recruitment from abroad). German and Spanish (but not French) are subjects in which students are less likely than others to be taught by a subject specialist. Weak evidence from stakeholders and inferences from official figures may also support the claim that there is a shortage of potential modern foreign language teachers coming through the English education system.

However, it is not clear whether the declining entries for GCSE French and German are in fact the result of severe grading (real or perceived) of these subjects. Uptake at GCSE has declined significantly, but from a starting point where modern foreign languages were compulsory at Key Stage 4. The decline in uptake at A level began (and a significant proportion of it occurred) while the subject was still compulsory at GCSE. Research also shows that perceived difficulty is not the most significant factor in students' decisions about which subjects they choose to study. They are more likely to choose to study a subject based on perceived enjoyment or utility. Taken together these points potentially suggest an alternative explanation for the decline, in the form of diminishing student interest. There are similarly declining entries in French and German in other parts of the UK and in Anglophone countries.

This makes it less convincing that grading standards are necessarily the most significant cause of declining entries.

Nonetheless, we considered the evidence of declining entry in French and German to be sufficient to regard this criteria as having been met. We were not satisfied that this was the case for Spanish, where entries are increasing. Other generalised evidence of potential issues we considered to be too weak to fulfil this criterion in Spanish as it was applicable to all three languages and did not provide a plausible explanation for the difference in entry trends between them.

Criterion C – Views of Users and Awarders

For our third criterion we considered evidence from users of the qualification, and the views of exam boards and their awarders on the acceptability of adjusting grading standards.

One approach we took to considering this issue was to benchmark these qualifications to the Common European Framework of Reference for Languages (CEFR) (for the first time as far as we are aware) through a linking study. Subsequently, stakeholders were surveyed on their expectations of performance at GCSE in terms of the CEFR levels to help establish whether the current standards demonstrated by students achieving grades 7 and 4, as benchmarked to the CEFR, represent an appropriate level of performance.

The linking study found that performance standards between the three languages were reasonably well aligned, with grade 4 at level A1 of the CEFR scale (although slightly higher performance was demonstrated in Spanish than French or German), grade 7 at mid-A2 level and grade 9 around low B1 for all languages. In the survey, the majority stakeholder view of the current performance standards at grades 7 (A2) and 4 (A1) broadly aligned with the levels identified in the CEFR benchmarking study, although awarders, possibly due to little familiarity with using the CEFR scale, appeared to overestimate current candidate attainment in relation to the CEFR. However, the majority of stakeholder responses, including those of awarders, also suggest that the current performance standards are lower than they thought they might ideally want them to be when considered against the CEFR scale. This is perhaps unsurprising given respondents were given the opportunity to express a view on where they would wish standards to be. It may also be that this tension reflects stakeholder preference to see some change to the focus of aspects of the assessment, and/or the focus of curriculum content. We have passed this evidence to the Department for Education to consider.

The results of the CEFR benchmarking study are also significant in relation to our duty to have regard to international comparability of qualifications standards. The study showed that the current performance standard at the lowest judgemental grade in each subject (grade 4) still registers comfortably on the CEFR scale,

suggesting that there may be scope for a cautious relaxation of grading standards without changing the broad CEFR levels which describe performance at key GCSE grades.

We also sought the views of exam boards, and asked their awarders about the acceptability of adjusting grading standards through a separate questionnaire. Two exam boards gave us their views. One felt the current grading standards are appropriate. The other argued they should be lowered as other attempts to address perceived severe grading through changes in qualification design and attempts to improve students' experience of taking these qualifications had failed to increase entries. Their awarders also disagreed along similar lines. Some of the awarders thought it would be acceptable to adjust grading standards in German; more, although not all, thought the same for French and Spanish, although they did not agree on the grades that should be adjusted. However, most awarders felt that relaxing standards might help increase uptake – even those who considered the current grading standards to be correct.

While the evidence under this criterion was mixed, overall we are of the view that there is sufficient weak evidence in support of a potential adjustment to grading standards in all three subjects to conclude that users and awarders of these qualifications would be likely to accept an appropriate relaxation of standards at key grades.

Criterion D - Potential Benefit vs. Wider Impact

Finally we considered whether the benefits of changing grading standards would outweigh any negative effects. We noted the broad stakeholder support for an adjustment across the sector, including subject associations, school leaders, and higher education. The strength of feeling from many of these stakeholders about the necessity of an adjustment was particularly strong. Some of these stakeholders suggested that the case for an adjustment was greater in French and German than in Spanish.

Exam boards and their awarders raised some concerns about potential negative impacts on progression or effective discrimination between different students if standards were changed. However, we are satisfied that an appropriate adjustment would be unlikely to create issues for higher education or compromise any of the stated purposes of GCSEs. In fact, the qualifications might provide a more effective basis for school accountability, one of the stated purposes of GCSEs, if grading standards are adjusted, so long as any adjustment was proportionate.

However, statistical modelling of the impact of aligning standards on grade boundaries suggested that there could be a notable drop in performance at certain grades if we aligned to a statistical mean of all GCSE subjects. This would be much more pronounced in French and German than in Spanish. This would have an impact on the performance standard and therefore the interpretation of GCSE

grades, in a context in which awarders may already be overestimating levels of students' performance, and a very significant adjustment could lead to altering the alignment with the CEFR we have observed. A change in this alignment resulting from lower performance would not be supported by our survey of stakeholders' expectations. Some stakeholders also noted potential negative impacts of an adjustment on other aspects of the education system, although these were weighed against the potential threat to public confidence of the perception that French, German and Spanish are more harshly graded than other GCSEs.

Overall we concluded that this criterion was met for all languages so long as any adjustment was appropriately calibrated.

Decision

We have judged the evidence presents a compelling case to adjust grading standards in GCSE French and German, where all four of our criteria were met, but not in Spanish, where we judged only two out of our criteria were fulfilled.

Adjustment to grade standards

We have decided to bring French and German grading standards in to line with those for Spanish. This is as far as we judge the evidence supports a negative impact on entries for modern foreign languages resulting from unduly severe grading. It also represents the furthest extent to which we can justify an adjustment given that we do not see this impact in (the statistically less severely graded) Spanish, where entries are increasing. We think it is also more justifiable than attempting to bring the standards closer to the mean of all subjects, given the volatility of that measure.

We have also chosen this specific adjustment because it aligns with the concept of shared minimum performance standards within languages, which share common subject content. We believe bringing standards for GCSE French and German in line with those for Spanish will not have an undue impact on performance standards established in these subjects.

As we have no statistical evidence and few specific stakeholder concerns about the relative difficulty of grades below grade 4, we have determined there is not a case to adjust grade 1 in any of the three subjects, but changes made at grade 4 could have a small impact on grades 2 and 3.

Implementation

We are considering with exam boards how best to implement this adjustment.

We will provide more information ahead of awarding these qualifications next summer. We will ensure the adjustment to align grading standards in French and

German with those in Spanish takes effect from 2020 (potentially phased over more than one year). It will mean that standards at all grades in GCSE French and German above grade 4 will either be made less severe or stay the same. We do not intend to change grading standards at grade 1, but changes made at grade 4 could have a small impact on grades 2 and 3.

This concludes our planned work on inter-subject comparability. However, we will keep the evidence in this area under review.

Introduction

This document summarises the evidence we have considered in relation to our investigation into grading standards in GCSEs in French, German and Spanish, and the criteria we used to decide whether there is a compelling case to adjust grading standards in these subjects. We set out our conclusions and explain the rationale for our decisions.

The investigation concludes our work to consider concerns from stakeholders that some GCSEs and A levels are more severely graded than others. We decided in April 2017¹ not to seek to achieve greater inter-subject comparability on a statistical basis between all GCSE and A level subjects. Instead we decided to investigate whether to adjust grading standards in some subjects only.

We announced the [outcome of our investigation into grading standards in A level science and languages](#) in November 2018. After analysing an extensive base of statistical evidence and contextual data, and considering a wide range of other evidence including detailed representations from the subject communities, we determined that there was not a compelling case to adjust grading standards. However, recognising the potential for perceived grading severity to undermine public confidence, we also committed to working with the exam boards to ensure these qualifications did not become statistically more severely graded in future.

We have used the same criteria as we used for these A level subjects when deciding whether to adjust grading standards for GCSEs in French, German and Spanish.

Our evidence

We considered evidence from a range of sources, including statistical evidence and contextual data, relevant academic papers and research. We also considered the quality of students' performance by looking at how GCSE performance and assessment standards at specific grades relate to the Common European Framework of Reference for Languages (CEFR), and conducted a stakeholder survey on their expectations of performance at GCSE in terms of the CEFR levels. We consulted with stakeholders including subject associations, Higher Education, headteachers' associations, teaching unions, cultural institutes, senior examiners and the exam boards. Where relevant, we have taken into account evidence previously gathered during our investigation into grading standards in A level modern foreign languages, although we have used more recent evidence where available.

We have included greater detail about the statistical evidence, along with entries data and the findings from our study benchmarking GCSE standards against the Common European Framework of Reference for Languages and the associated

¹ [A Policy Position for Ofqual on Inter-subject Comparability](#)

survey in accompanying annexes. We have summarised these sources and the other relevant evidence we have considered in this document. Full details of this other evidence (for example, academic papers and reports produced by organisations other than Ofqual) are provided in the references section.

A summary of the criteria

When we considered grading standards in A level sciences and modern foreign languages, we devised a set of criteria that could be applied to any subject to consider whether there was a compelling case to adjust grading standards. We have used the same criteria in our investigation into grading standards in GCSEs in French, German and Spanish, with some minor changes to reflect the different context of these qualifications from A levels.

The criteria reflect the view that no single piece of evidence will definitively demonstrate the case for an adjustment to grading standards in a given subject (particularly in light of the limitations identified in the statistical evidence). Any decisions will ultimately be judgements based on an holistic consideration of a wide range of factors. The criteria are set out below:

- A. Statistical measures of relative subject difficulty show evidence of potential misalignment over several years. If this is the case we would expect to see evidence of the following:
 - i. Different forms of statistical evidence from several years of entry align to indicate the subject is consistently more difficult than most other subjects
 - ii. The average level of difficulty of the qualification, as indicated by statistical measures, is substantially above the average of key and/or most grades

- B. Persuasive evidence of the potential detrimental impact caused by perceived severe grading on those who use the qualification and on society at large. If this is the case we would expect to see persuasive evidence of negative impacts, which might include the following:
 - i. Depressed uptake of courses to which students taking the subject would be expected to progress
 - ii. Depressed uptake of the qualification

- iii. Indications of issues in securing a sufficient supply of teachers
 - iv. Indications of skills shortages related to a lack of take up of the subject
- C. Evidence of dissatisfaction with the current grading standard from those who use the qualification, and those responsible for maintaining the standard.
Under this criterion we would consider:
- i. The views of those who use the qualification
 - ii. The views of the exam boards, and specifically the judgements of examiners responsible for making awarding decisions
- D. The likely benefit to users of the qualification and society as a whole from a change to grading standards must outweigh any potential negative effects. To judge if this the case, we would expect that:
- i. There is evidence of support from users of the qualification for any change
 - ii. There is no reason to believe that there would be a detrimental impact on the extent to which the subject fulfils the defined purposes of the qualification
 - iii. There is no reason to believe any change would have a detrimental impact on performance standards, for example by decreasing the level of cognitive demand in comparison to other cognate subjects
 - iv. There is no reason to believe that there would be a significant detrimental impact to other parts of the education system as a result of an adjustment

The evidence under Criterion A

Criterion A:

Statistical measures of relative subject difficulty show evidence of potential misalignment over several years

We use terms such as ‘severe’ and ‘lenient’ in reference to the apparent difficulty of the GCSE subjects in question under statistical measures rather than a judgment of whether they are actually more severe or lenient overall. The same applies to the use of the terms ‘easy’ and ‘hard’ when used in relation to statistical measures of subject difficulty. When we use such terms we are considering the grades a student achieves in the subject compared to the grades they achieve in other subjects.

i. Different forms of statistical evidence from several years of entry align to indicate that the subject is consistently more difficult than most other subjects

Does the statistical evidence indicate that GCSE French, German and Spanish are statistically more difficult than other subjects?

We discussed the limitations of statistical measures of subject difficulty such as Rasch analysis in our work to consider the case for an adjustment to 6 A level subjects in 2018 ([see Black et al., 2018](#)). Most of these measures are conceptually reliant upon the existence of a linking construct, such as an underlying ‘generic academic aptitude’. This presumes that a student should perform just as well in an assessment in music as they would in English, or in history as they would in physics. It ignores any potential subject-specific effects such as intrinsic demand, effectiveness of teaching, allocated curriculum time, personal motivation, or individual aptitude. We therefore decided that that the evidence under this criterion would need to be corroborated from multiple sources for us to consider it to be persuasive.² There can also be issues with the statistical methods generating results which can be unreliable, or artefacts of the data analysed (for example, violation of model assumptions by real data, imperfect data-model fit, and non-random missing data resulting from student subject choice). Even where different measures agree, we should treat statistical evidence with caution.

We present the results from a range of statistical analyses (including Rasch analysis, subject pairs analysis, Kelly’s method, reference tests and value-added models) in

² This is consistent with our approach at A level.

our statistical evidence report (He & Black (2019)). These analyses suggest that GCSE French, German and Spanish are generally more “statistically difficult” than the average difficulty of all GCSE subjects at both individual grade level and overall subject level. When the same analyses are carried out on National Pupil Database (NPD) data from 2002 – 2018, it appears that this has consistently been the case for approximately the past 15 years. In 2006, all three subjects appeared to be at least two fifths of a grade more difficult than the mean of all subjects based on the Rasch model, and slightly less than a third of a grade harder based on prior attainment models. However, while statistical analysis can provide useful information about the relationship in performance between different subjects, caution must be exercised when interpreting any subject difficulty measures derived and linking these difficulty measures directly to the grading standards of examinations which are generally subject specific.

If we imagine a spectrum of difficulty across all GCSE subjects, Rasch analysis suggests that in 2016, French, German and Spanish were amongst the five hardest of the 30 significant entry GCSEs, along with Latin (which appeared most difficult) and Business Studies. Repeating this analysis on 2019 summer awarding data shows that the overall relative difficulty ranking of the GCSE subjects had changed³, with Spanish now appearing to be the seventh most difficult subject but French and German still among the five hardest GCSEs.

Statistical measures indicate that there are differences in the apparent level of difficulty between the three languages. While they may all be of above average difficulty compared to all GCSE subjects (when considered at overall subject level, rather than individual grade level, at least), certain languages appear to be more severely or leniently graded than others in terms of the grades students achieve in them when considered alongside their achievement in other subjects. This is despite the fact that GCSEs in French, German and Spanish all share common aspects of subject content and rules around assessment design, possibly pointing to inherent differences in the difficulty for English students of learning each language rather than a misalignment of grading standards. In 2016, French and Spanish were about a third of a grade more difficult than the mean of all subjects, while German was about two fifths of a grade harder.

³ This routine year-on-year fluctuation in the relative difficulty of subjects under statistical measures, such as Rasch, is an important point to note when considering an adjustment to grading standards, as it suggests that a one-off adjustment may not guarantee a change in apparent relative difficulty ‘rankings’ of this kind if the difficulty of other subjects continued to fluctuate in the future.

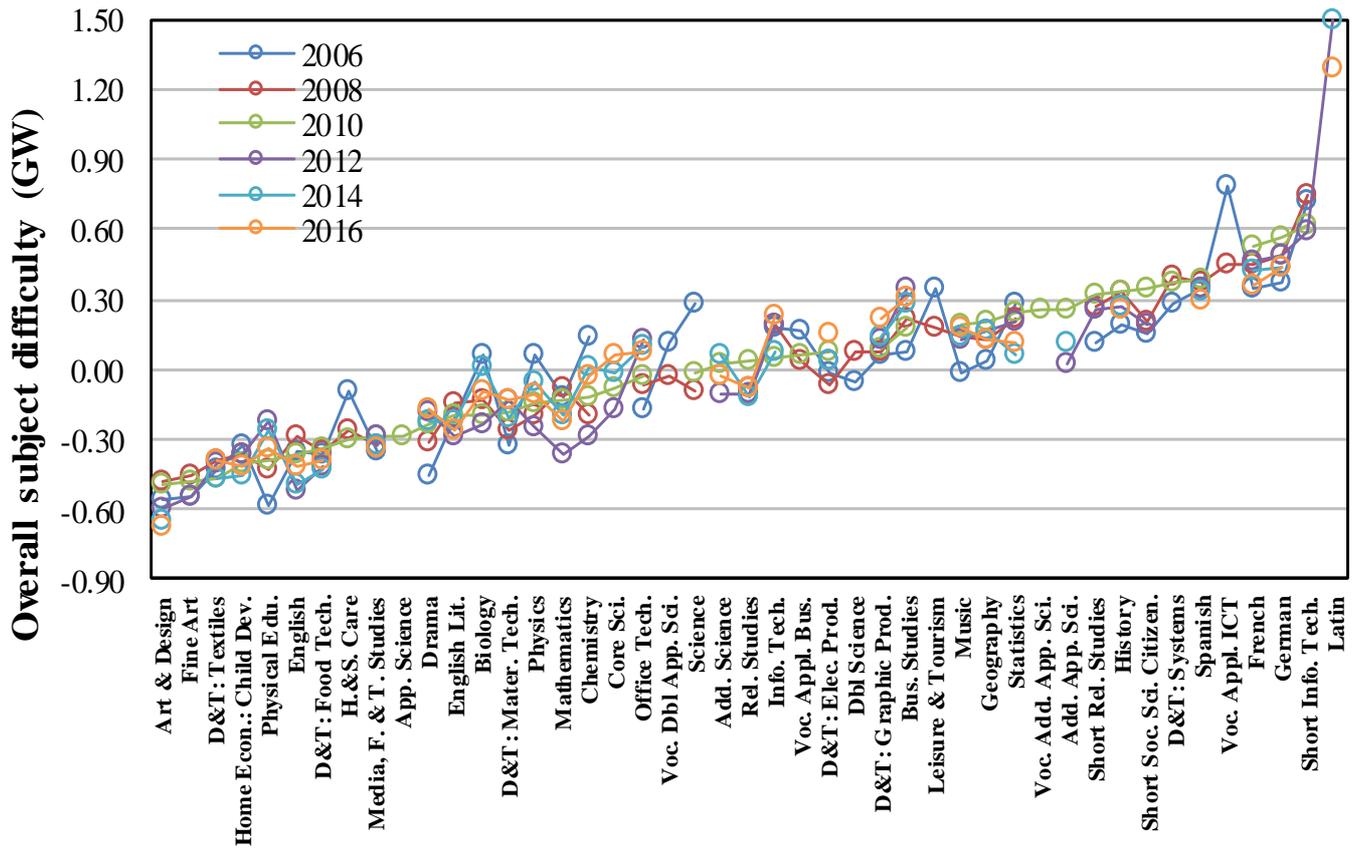


Figure 1: Distribution of the overall subject difficulty (in grade width) of GCSE subjects from 2006, 2008, 2010, 2012, 2014 and 2016 based on NPD data and using the Rasch model (subjects are ordered by overall difficulty in 2010). This illustrates the volatility of relative difficulty measures year to year.

However, this difference in difficulty is not consistent between different years of entry and under different statistical measures. For instance, when considering relative outcomes in all GCSE subjects in 2010 in relation to prior attainment at KS2, French and German appear to be about a quarter of a grade harder than the average of all subjects, and Spanish only about a tenth of a grade harder. When repeated for prior KS2 prior attainment and results for the cohort taking GCSEs in 2016, outcomes in French, German and Spanish were still lower than those expected based on their KS2 profiles, but the difference in difficulty between these subjects and GCSE geography and history was generally smaller than that in 2010. Indeed, in 2016 Spanish was of the same apparent difficulty as GCSE history according to this measure. Under nearly all of the statistical analyses GCSE Spanish is closer to the statistical mean of all other subjects than French and German (while still appearing harder than the average difficulty of all subjects).

The approach to maintaining standards at awarding which we have adopted since 2010, and our decision to prioritise preserving grading standards between the legacy and reformed GCSEs, means that although the apparent statistical difficulty may

fluctuate slightly the actual difficulty of these subjects will not have changed appreciably, despite changes in subject content and assessment demand. Value added analysis, looking at relative performance in the reformed GCSEs in 2018 (the first year of examination) by KS2 prior attainment, showed that students with middle to high levels of prior attainment generally achieved slightly lower outcomes in GCSE French, German and Spanish than in other subjects included in the analysis, again suggesting that they were more difficult in statistical terms.

Overall the majority of the statistical analyses presented in our statistical evidence report (He & Black (2019)), and from published studies over the past two decades, suggest modern foreign languages, and particularly French and German, are more difficult than most other GCSEs in statistical terms. We might also place greater weight on the statistical evidence when considering the advisability of an adjustment to grading standards at GCSE than we did at A level. Some of the limitations of Rasch and other statistical measures of subject difficulty identified by Bramley (2016), where the existence of non-random missing data as a result of the combinations in which students are likely to select A level options was shown to produce biased estimates of subject difficulty, are likely to be lessened (although not eliminated) by the greater proportion of students taking these subjects at GCSE and the fact that certain subjects (such as English and mathematics) must be taken by all at KS4.

Is this proof that the standard in GCSE French, German and Spanish is severe?

Statistical analyses can provide useful evidence about the relationship in performance between different subjects. However, as noted above, there are limitations to this source of evidence when it is used to compare grading standards. For most of the statistical methods considered above, it is explicitly or implicitly assumed that examinations in different subjects share a common construct or latent trait which is closely related to the constructs being measured. The difficulty of a subject is normally defined as the amount of the common trait required to achieve a specific level of performance in the exam. The statistically defined difference in difficulty between subjects may reflect variation in the amount of this common trait required to achieve the same level of performance. However, it does not necessarily imply that some subjects are graded more severely (or leniently) than others. This is because performances at GCSEs and A level are also graded according to the judgement of subject expert awarders based on standards representing an expectation of knowledge and skills which are subject specific. This approach to awarding does not assume that students of a given prior attainment *should* have an equal probability of gaining a specific GCSE grade across all subjects. It might be entirely legitimate that a student of a given level of prior attainment performs to different standards in different subjects because of factors such as teaching time and prior exposure.

Performance standards are maintained year-on-year using statistical evidence, matched prior attainment and senior examiner judgement. If a cohort's prior attainment is similar to that of the cohorts in previous years, we would expect the proportion of students achieving each grade to be similar. We require exam boards to use senior examiners' expert judgement to check the grade boundaries that these statistics indicate, and provide evidence to us in cases where they would propose outcomes should be greater or lower than the statistical evidence would indicate is appropriate. This approach has meant that GCSE results have generally stayed stable since 2010, but statistical measures show that the relative difficulty of GCSEs in French, German and Spanish, when compared to other subjects, has fluctuated over the same period. These patterns of variability are broadly similar in all three subjects – their difficulty increased slightly from 2006, peaked in 2010, and decreased gradually from 2010 to 2016. Spanish shows less variability than French or German overall in that it appears to have generally remained closer to the statistical 'average' difficulty of all GCSEs, although it does show greater variability at individual grades under some measures. These fluctuations, in spite of stability in GCSE outcomes, suggest that there are other factors which may be contributing to the relative difficulty of these subjects beyond the grading standard being applied and indicates that this evidence must be treated with a degree of caution.

ii. The average level of difficulty of the qualification, as indicated by statistical measures, is substantially above average at key and/or most grades

Is this apparent statistical misalignment consistent across all grades in GCSE French, German and Spanish, or only present at some grades?

Statistical evidence clearly indicates variation in relative subject difficulty for different grades within each of GCSE French, German and Spanish. This suggests that they are not all equally difficult at all grades in statistical terms – or 'difficult' at all, in cases where statistical difficulty dips below the average of all subjects. Furthermore, patterns vary between each of the three languages, and over time. Modern foreign languages as a whole are not consistently more difficult at all grades than other GCSE subjects, nor are they all consistently more difficult at the same grades. The picture of relative difficulty indicated by statistics is more complex and changeable.

Our statistical evidence report (He & Black (2019)) includes Rasch modelling of relative subject difficulty at grades A*, A and C for GCSE French, German and Spanish for the period 2002 – 2016. This predates the introduction of 9 to 1 grades in this subject, but direct grade equivalents in the reformed GCSEs are provided as appropriate. We have included both A* to G and 9 to 1 graded GCSEs in our analysis as this allows us to consider the relative statistical difficulty of these subjects

over an extended period of time. This is important given stakeholder arguments that these subjects have been affected by a long-standing misalignment of grading standards, which has resulted in them being more severely graded than other GCSEs over a period which coincides with a significant decline in entry. As grading standards were maintained at the point of reform, the statistical relationships outlined below have been carried forward into the new qualifications.

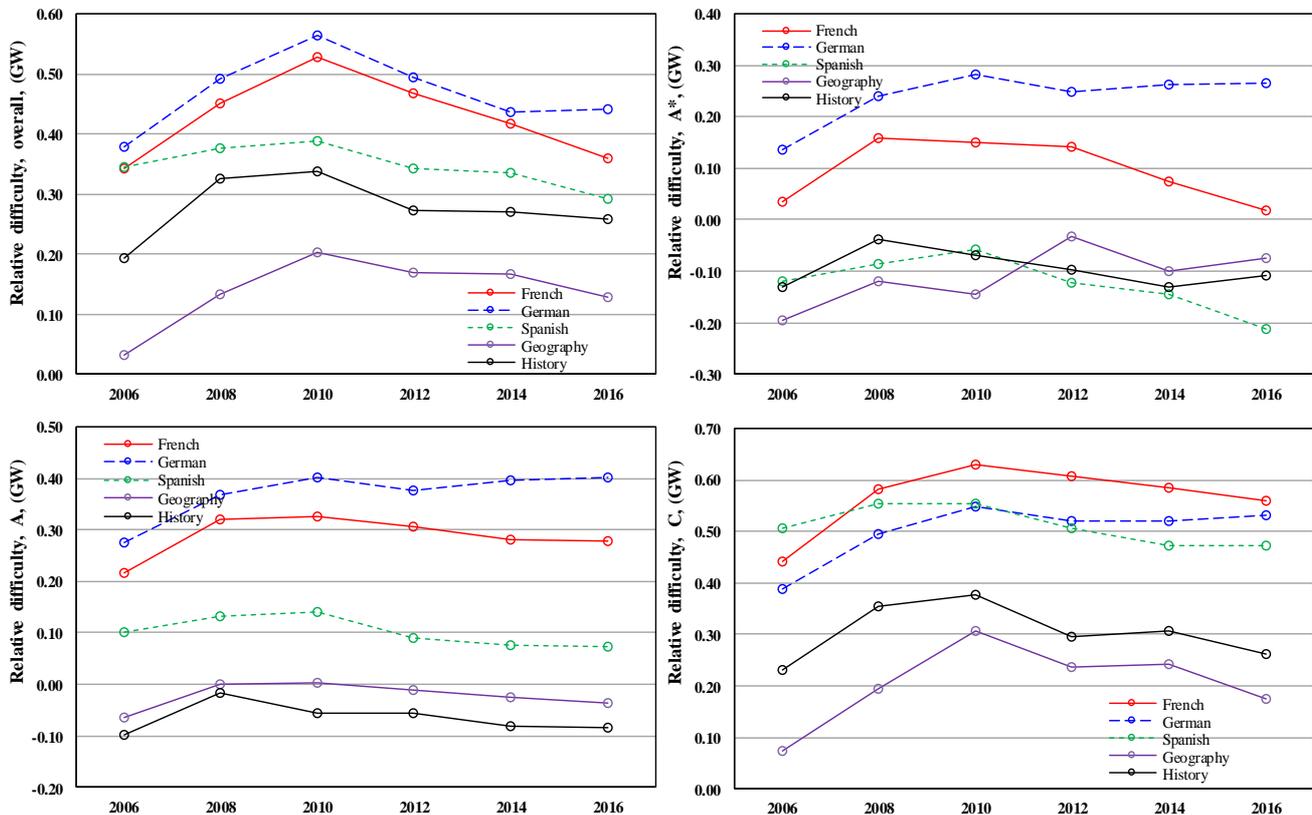


Figure 2: Distribution of overall subject difficulty and difficulties at A*, A and C for GCSE French, German, Spanish, geography and history from 2002 to 2016 based on NPD data and using the Rasch model.

At A*, both French and German were harder overall than the average of all subjects but Spanish was easier. In recent years the difficulty of A* in French has been settling at a point reasonably close to the average for all GCSEs. The relative difficulty of German went up slightly from a tenth of a grade higher than the average difficulty of all subjects in 2006 to nearly a third of a grade higher in 2010. It then decreased slightly from 2010 to 2012, and remained about a quarter of a grade higher than the average from 2012 to 2016. For French, the relative difficulty in 2006 was close to the average of all subjects. It then went up slightly to about a sixth of a grade higher in 2008, and decreased gradually from 2008 to the average difficulty in 2016. The relative difficulty of Spanish was about a tenth of a grade lower than the average in 2006. It then went up very slightly in 2010, and decreased gradually from 2010 to nearly a fifth of a grade lower than the average in 2016.

At grade A/7, German was harder than both French and Spanish. It was also about a quarter of a grade harder than the average of all subjects in 2006. Its relative difficulty increased to about two fifths of a grade higher than the average difficulty in 2010 and remained at that level since. French was about a fifth of a grade harder than the average in 2006. Its difficulty increased to nearly a third of grade harder than the average in 2010, but decreased gradually from 2010 to slightly over a quarter of a grade higher than the average in 2016. Spanish was about a tenth of a grade harder than the average in 2006. Its relative difficulty went up slightly to just over a tenth of a grade harder than the average in 2010 and decreased very slightly from 2010 to below a tenth of a grade harder than the average in 2016.

At grade C/4, French, German and Spanish were all harder than the average of all subjects. French was slightly over two fifths of a grade harder than the average of all subjects in 2006. Its relative difficulty went up to slightly to over three fifths of a grade above the average in 2010 and decreased gradually from 2010 to around half a grade above the average in 2016. German was nearly two fifths of a grade harder than the average in 2006. Its difficulty went up to around half a grade above the average in 2010 and has remained broadly at this level since. Spanish was about half of a grade harder than the average in 2006. Its difficulty increased slightly in 2010 and decreased gradually from 2010 to below half of a grade above the average in 2016.

Some stakeholders believe that a significant proportion of native speakers are taking these subjects, systematically influencing awarders' expectations in a way which makes it more difficult for the rest of the cohort to achieve higher grades in all three subjects. The evidence does not support the view that there is a greater statistical difficulty at grades 7, 8 and 9 across all three subjects. At C/4 certain statistical analyses suggest that these subjects are approximately 0.5 grade widths harder than the average of all GCSEs. On the basis of this evidence, the case for adjusting grading standards at grades 7, 8 and 9 (covering grades A and A* in the legacy qualifications) is stronger in French and German than Spanish. There is not a statistical case for adjusting standards at grade 1 in any subject.

This is illustrated by the modelling presented in our statistical evidence report (He & Black (2019)) of the changes that would be required to grade boundaries in GCSE French, German and Spanish to align them to the statistical average of all subjects based on Rasch analysis.

Conclusion

The evidence from statistical measures of subject difficulty suggests that GCSEs in French and German are more severely graded than the majority of other GCSEs at most grades. This appears to have been consistently the case for a significant period of time. There is a lack of consensus about what these statistical measures actually

indicate, and strong arguments about why there might be legitimate differences in the relative demand of different subjects. However, we believe this evidence is sufficient to suggest that grading standards in French and German differ enough statistically from those in other GCSEs used for similar purposes in schools to fulfil our criterion.

The evidence is not the same for GCSE Spanish however, and as such we have not judged this criterion to be fulfilled. This subject does not appear to be systematically more difficult than other GCSEs at all grades, and in fact at some grades appears less severely graded than the statistical average. The difficulty of Spanish also appears to fluctuate more by year. Overall the subject appears to have become relatively easier in statistical terms in recent years. It also appears under these measures to be easier than French and German, and more closely aligned to (and at some grades easier than) other optional EBacc subjects, for example GCSE history and geography.

It is also important to note the lack of precision that would be inherent in any adjustment based on statistical measures of subject difficulty alone. The evidence in our statistical evidence report (He & Black (2019)) shows that the relative difficulty of many subjects varies year-to-year, likely a result of routine factors such as changing entry patterns. This is another illustration of why we cannot rely solely on statistics in judging the case for an adjustment – as it demonstrates that the relative position of subjects according to Rasch analysis and other statistical measures can change even when there is no material change in the difficulty of an individual subject. It also illustrates why there is a strong case that any adjustment to grading standards should be made on a one-off basis.

The evidence under Criterion B

Criterion B:

Persuasive evidence of the potential detrimental impact caused by perceived severe grading on those who use the qualification and on society at large

i. Depressed uptake of courses to which students taking the subject would be expected to progress

What is the relationship between apparent subject difficulty and progression to A level French, German and Spanish study?

Entries for A level French and German have declined markedly since 2002, although figures presented by Macaro (2008) demonstrate that this is in fact a trend which began as far back as 1992. A level French, German and Spanish entries generally follow a similar pattern to that seen at GCSE. In 2018, A level French entries were almost equal to Spanish entries, compared to 2012 when French entries were almost three times that of Spanish. The decline has been more severe in French than German, although entries for French were significantly higher to begin with and the proportional decrease has been similar. In contrast, entries for Spanish have risen significantly over the same period, although the rate of increase has slowed recently.

The expectation that students must have studied the language at GCSE to be able to progress to the corresponding A level means a decrease in GCSE entries will result in fewer potential A level candidates. Rates of progression in French, German and Spanish from GCSE to A level have remained relatively stable, but are lower than those seen in other optional GCSE subjects such as history and geography.

The proportion of students who progress from GCSE to A level has been generally steady, despite significant changes in entry numbers, suggesting that if GCSE entries were to increase then so would those at A level. However, research by Macaro (2008) does challenge the assumption that increased uptake at GCSE would necessarily lead to more students choosing to study modern foreign languages at A level. Macaro notes that the most significant proportion of the reduction in A level French and German entries took place in the period 1992 – 2002, during which time modern foreign languages were compulsory at KS4. In other words, at the point at which the adoption of a 'Languages for All' policy meant there was the greatest possible pool of potential A level linguists, this was not reflected in the number of students who progressed to study the subject at A level. In fact compulsory study could have had the opposite effect, actively putting some students off. This

corresponds with the findings of Parish and Lanvers (2018) which concluded that choice was linked to higher intrinsic motivation in language learning, but also that choice was only beneficial to motivation when students had a completely free choice.

This suggests a more complex relationship between student motivation and the decisions made by schools about access to GCSE modern foreign languages, which may question whether, if all other factors remain the same, increases in the number of students studying GCSE French or German would result in significant increases in the numbers taking those subjects at A level – conflicting with the evidence of GCSE to A level progression rates cited above. Overall, we are of the view that an increase in entries for GCSE French, German and Spanish is more likely to lead to increased entries at A level than not. However, it may be the case that the impact of such an increase in entries at GCSE on A level progression would be lower than some might expect.

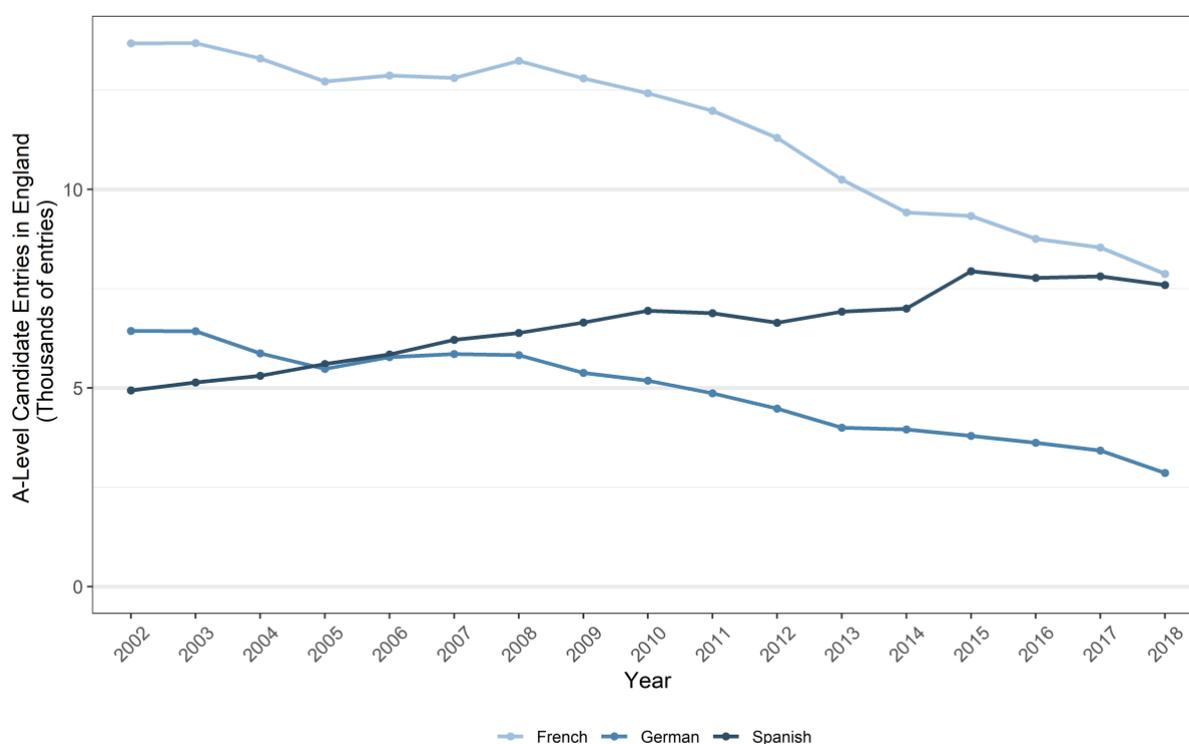


Figure 3: Number of entries for French, German and Spanish A level exams between 2002 and 2018 in England. Data extracted from JCQ A level main results tables for summers 2002-2018.

The decline in modern foreign languages entries could also be seen as an inevitable consequence of increased uptake of STEM subjects (following efforts to increase their uptake). Indeed, A level sciences and maths have seen significant growth in recent years at the expense of other subjects traditionally regarded as ‘facilitating’ entry into higher education⁴. A level English, for instance, has also experienced a

⁴ The definition of ‘facilitating subjects’ has now been withdrawn by the Russell Group of universities.

significant decline in entries in recent years, and yet Rasch analysis based on the 2017 NPD showed that A level English was statistically easier than the average of all subjects – and much more so than physics, chemistry or biology, which all saw a substantial increase in entries over the same period.

Nor did our previous investigation into A level modern foreign language grading standards find compelling evidence that perceived difficulty based on outcomes was likely to be the primary motivating factor in students' decision to opt to study a subject. Instead, students' decisions were found to be based on a combination of factors which prioritised personal enjoyment/interest and likely future career utility. While it could be that some students are dissuaded from studying A level modern foreign languages because they find GCSE to be 'difficult', there is a broad range of factors which may contribute to experiential difficulty which are distinct from grading standards. Responses to the British Council's *Languages Survey 2019* placed 'the nature and content of external exams' above their marking and grading in both state and independent schools as the key factor affecting take-up at GCSE and A level. The report also identified a range of other factors which are likely to directly contribute to students finding languages to be 'harder' than other subjects, including:

- variable languages provision at KS2
- lost progress on transition to KS3
- relatively fewer timetabled teaching hours in comparison to other subjects
- declining status of languages in the eyes of school leaders and parents negatively impacting on student motivation
- loss of enrichment activities such as pupil exchanges and with them opportunities to practise languages being learnt
- issues with the recruitment of suitably qualified teachers

These are also factors which could have an impact on student progress in these subjects and potentially lead them to appear to be more 'difficult' under statistical measures than other GCSEs.

ii. Depressed uptake of the subject

What is the relationship between apparent subject difficulty and uptake in GCSE French, German and Spanish?

Language stakeholders point to a long-term decline in entries for GCSE modern foreign languages as evidence of the impact of severe grading, and as a significant issue in terms of the 'pipeline' of students who may progress to study the subject at A level and in Higher Education. Entry figures for GCSE French, German and Spanish in England show that this is not the case across all subjects however.

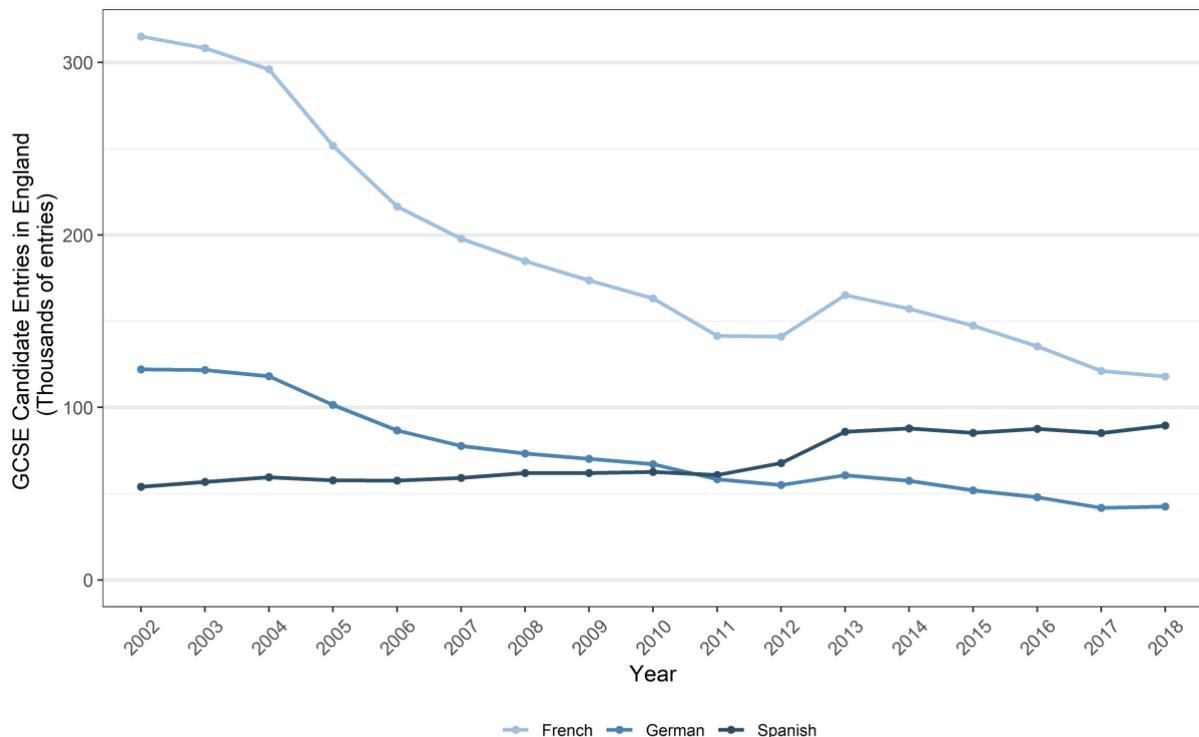


Figure 4.: Number of entries for French, German and Spanish GCSE exams between 2002 and 2018 in England. Data extracted from JCQ GCSE main results tables for summers 2002-2018.

Since 2002, Spanish GCSE entries have been increasing, to the point where they overtook German GCSE entries in 2011. They have continued to increase and in 2018 had only 30,000 fewer entries than French GCSE.⁵ However, there has been a significant reduction in French GCSE entries which, since 2002, have declined by almost two thirds from 315,000 to 118,000. The beginning of this trend coincides with the point at which languages ceased to be compulsory in England at KS4. From 2012 to 2013 all three MFL subjects saw an increase in entries (potentially linked to

⁵ This gap continues to narrow: 2019 provisional entry figures projected an increase in GCSE French entries from 120,605 to 125,700 and suggested GCSE Spanish entries would also rise from 91,980 to 101,080. While this also marks a change in the general trend for French entries, GCSE German entries appeared set to fall again slightly.

the inclusion of modern foreign languages within the government's EBacc performance measure) which was most pronounced in French. However, following this French entries have again declined, in comparison to Spanish which continues to grow after a period of relative stability and German which may now also be stabilising.

It is not a given that students base their decisions about GCSE study options based primarily on their expectation of the grade that they will achieve. Indeed the available evidence suggests that this is just one factor among many which informs their decision and also that it is frequently of lesser importance than their enjoyment of the subject or the extent to which they feel that learning a modern foreign language will contribute to their future career prospects.

Department for Education (DfE) research into subject choice motivations carried out via the [Omnibus Survey of Students and their Parents or Carers](#) has consistently (see for example the report published in March 2019) shown that students generally rate enjoyment and utility for their chosen career as being a greater motivating factor than their perceived success in a subject when choosing options for study.

Subsequent survey waves indicate that 50% of students surveyed did not have a choice over whether or not to study a language (either because it is compulsory, or because their school does not offer these subjects at GCSE); that parents and carers are strongly supportive of language study and their motivations for advising whether or not to study a language follow similar priorities to those of students; and that students' decisions about taking a modern foreign language are fixed at a relatively early point of Key Stage 3.

The findings of the DfE survey are supported by research by Cuff (2017) which indicated that subject choices were driven by three perceptions – enjoyment, usefulness and difficulty – with perceived difficulty having the least influence on students' decisions. Cuff also found that students' perceptions of difficulty were highly subjective, and only occasionally corresponded to broader conceptions of inter-subject comparability. However, teachers and school leaders are influenced by their perceptions of subject difficulty, and this could have an impact on students' decisions or their access to subjects which are viewed as being harder than others.

Taylor (2015) also found that students' subject choices were more likely to be influenced by beliefs about the positive outcomes associated with a particular subject (for example, future career prospects) than they were by beliefs about difficulty. Furthermore, parents' and carers' views were found to be a stronger likely influence on students' decisions than those of their teachers.⁶

The hypothesis that factors other than grading severity may be leading students to choose to study other subjects at GCSE in preference to French, German or Spanish (in schools where that choice exists) is lent further weight when entry trends in

⁶ Although it is important to note that in the context of GCSE students, they may have no choice if languages are either compulsory for all at KS4 or simply not offered (two trends identified in the British Council report). So individual student motivation may play a less significant role than at A level.

England are considered in comparison to equivalent qualifications in the rest of the UK, and to other Anglophone countries such as New Zealand and the United States. A consistent decline in entries for French and German, coupled with an increase in entries for Spanish, can be observed not only in countries where grading standards for qualifications are designed to be aligned with those in England (such as Wales and Northern Ireland), but also in countries elsewhere where grading standards have never been aligned with GCSEs.

Indeed, the only Anglophone country considered where entries for French and German appeared to be relatively buoyant (with uptake for German actually increasing) was Ireland – where we understand that entry policies in schools usually make study of a modern foreign language for the junior certificate effectively compulsory. Even there figures showed a similar increase in subject entries for Spanish.

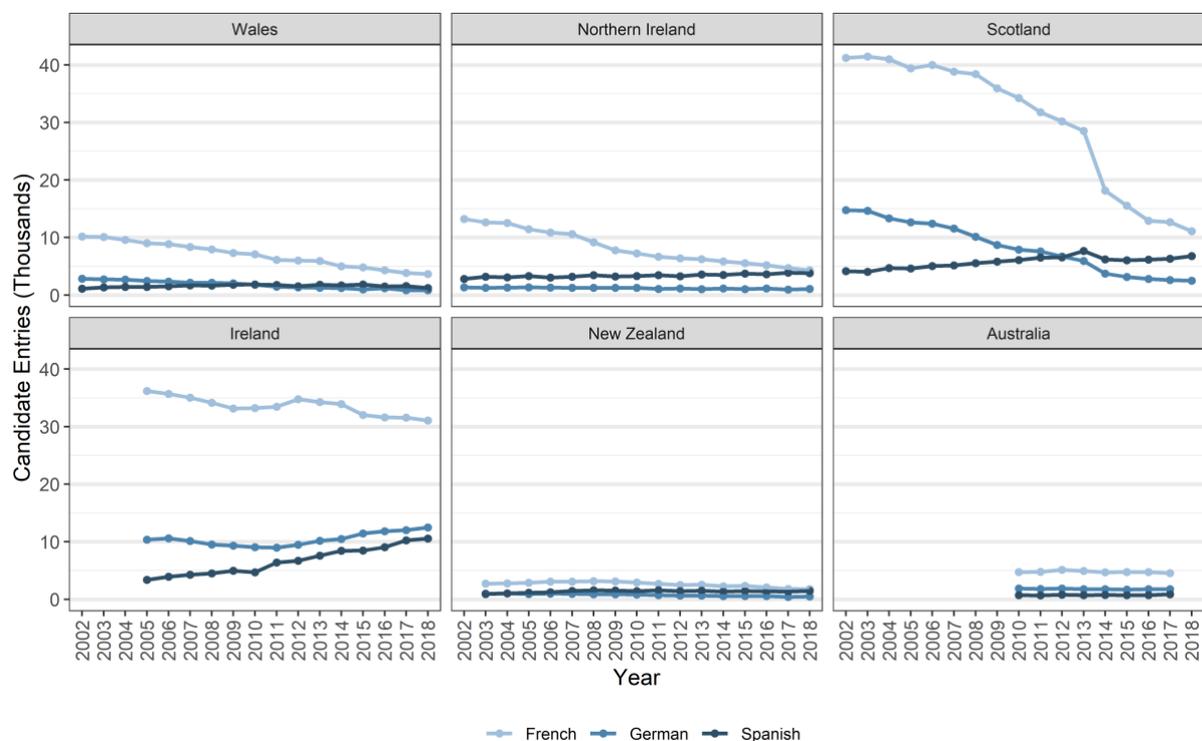


Figure 5: Side by side comparison of trends in modern foreign language entries across English speaking countries (excluding England). Note the graphs for New Zealand and Australia are condensed by the scale on the y axis which makes them appear more stable, but show similar trends to England.

When considered alongside research into what motivates students' subject choice, these trends call into question the cause and effect relationship between apparent severe grading in English GCSE modern foreign language qualifications and declining uptake. Logically, it also raises the question of whether an adjustment to grading standards would be likely to result in any meaningful improvement in entries for GCSE French or German.

iii. Indications of issues in securing a sufficient supply of teachers

Is there evidence to suggest that the perceived difficulty of modern foreign languages is leading to issues with teachers supply?

Definitive figures on teacher recruitment to modern foreign language subjects in England are lacking. The best indicators available suggest that issues with teacher supply in this subject area are becoming more acute, but do not provide us with a clear indication of the likely cause. There is also a potential ‘chicken and egg’ argument in relation to the perceived difficulty of GCSE French, German and Spanish and the supply of suitably qualified teachers. It is conceptually possible that apparent severe grading could be either a cause of recruitment issues or a consequence of this if students were less well taught.

DfE figures on recruitment to Initial Teacher Training (ITT) for the 2018/2019 academic year are forecasting a 12% shortfall in the number of new modern foreign language teachers necessary to meet the need identified under the Teacher Supply Model (TSM), compared to a 5% shortfall in 2015/2016. The number of awards of Qualified Teacher Status (QTS) made to teachers from overseas who qualified in the European Economic Area (a logical source of teachers for French, German and Spanish) also declined by 25% in 2017/2018, although the number of new ITT entrants from the European Economic Area (a measure which encompasses all subjects, not just modern foreign languages) has remained stable .

The Ofqual report into grading standards in A level modern foreign languages noted that the DfE believes that the main policy levers available to increase teacher recruitment have already been employed. It also noted that the National Audit Office (NAO) has provided alternative figures on teacher recruitment shortages which suggest that the situation is worse than reported by the DfE, and likely to become more severe in the future. Furthermore, the NAO identified Spanish and German as subjects which were particularly likely to be delivered by a non-specialist, with 43% and 25% respectively of lessons being taught by teachers with no relevant post A level qualifications in those subjects.

In addition, the British Council’s *Languages Trends 2019* survey found that 34% of state schools and 24% of independent schools surveyed reported difficulties in recruiting languages staff. These shortages are not evenly geographically distributed however, but appear particularly acute in certain areas – being more pronounced in the South than the Midlands and the North, and especially severe in the East of England, where 47% of schools were struggling to recruit.

iv. Indications of skills shortages related to a lack of take up of the subject

Is there evidence of skills shortages as a result of the declining take up of modern foreign languages?

The All Party Parliamentary Group (APPG) on Modern Languages and the British Academy have both cited concerns from employers about a lack of sufficiently competent linguists amongst the workforce. The *National Recovery Programme for Languages* proposed by the APPG argues that “the UK’s languages deficit is holding us back economically, socially and culturally”. The APPG claims that over 80% of Small to Medium Enterprises (SMEs) operate in English only, despite a reported 43% higher export/turnover ratio amongst SMEs using languages. The British Academy cites research estimating that a lack of language skills costs the UK £48bn a year, or 3.5% of GDP.

In secondary schools, the 2017 NAO report *Retaining and developing the teacher workforce* found that the number of teachers with a relevant post A level qualification in French, German or Spanish is increasing. However, while the proportion of French teachers possessing a post A level qualification was relatively high, the equivalent figures for German and particularly Spanish were lower.

DfE figures from the *Initial Teacher Training (ITT) Census* indicate that while the number of ITT entrants possessing a 1st class degree has increased in recent years, so too has the number holding a degree classification of 2:2 or below. This may suggest that while incentives, such as bursaries, are making language teaching a more attractive option to high performing graduates, a shortage of linguists amongst the pool of potential teachers is leading ITT programmes to relax entry requirements.

The British Council report also found a strong dependence in English schools upon languages teaching staff from overseas, particularly EU27 citizens. In some cases these constituted a significant number, or even all, of the teachers in modern foreign languages department in the schools surveyed. While this may be interpreted as further evidence of a shortage of sufficiently qualified British linguists, it is also presumably a career which EU27 citizens who are native speakers of modern foreign languages taught at GCSE and A level might be expected to pursue. However, when considered in light of the declining numbers of QTS awards to teachers who qualified elsewhere in Europe, this does suggest that a previously common source of modern foreign language teachers may be beginning to wane, which will likely compound the issues in ITT recruitment against the TSM identified above.

Overall then there is some evidence of skills shortages. Some stakeholders have attributed these shortages to severe grading, reflecting the impact on the flow of students taking languages to those becoming language professionals.

However, the nature of this evidence, and the fact that it generally fails to reflect the differences between individual subjects, reporting instead on the situation in modern foreign languages as whole, means that some of it is relatively weak.

Conclusion

There is clear evidence of a decline in entry for GCSE French and German (and a similar decline in entries for these subjects at A level), but this is not the case in Spanish. This is certainly strong evidence of an issue in uptake in French and German. The increasing shortfall in ITT entrants for languages and more general staffing issues may be indicative of a period in which recruitment may be particularly challenging, as suggested by some stakeholders.

However, some of the evidence of issues which stakeholders attribute to the impact of severe grading is in fact also a plausible alternative explanation for the perception that these subjects are harder than others – potentially accounting for their relative statistical and experiential difficulty. For instance, students whose teacher is not a languages specialist, does not hold a relevant post-16 qualification in that specific language, or students who do not have a regular teacher at all, might reasonably be expected to make less progress in the subject and find it harder in comparison to others. The British Council's *Language Trends 2019* survey report also identifies a number of other challenges – such as poorly handled transition from Key Stage 2; reduction in allocated teaching time in some schools, particularly in institutions with a truncated Key Stage 3; and fewer opportunities for students to practise languages outside of the classroom. This might have an impact on the attainment, motivation and progress of students, and in turn lead them to regard these subjects as more difficult than others. This offers a reason to be cautious about adjusting standards. Nonetheless, given that French and German have experienced a prolonged and significant decline in entries and evidence suggests that teacher recruitment is becoming more difficult generally in these subjects, we have determined this criterion is fulfilled in these subjects.

The position in Spanish is different. Spanish entries for the subject are increasing rather than declining at both GCSE and A level. Although some of the evidence suggests that a number of the issues identified above are common to all three languages, it does not provide a plausible explanation for the difference in entry trends between them and is therefore too weak to fulfil this criterion for Spanish. Given both these points, we have concluded that this criterion is not met for GCSE Spanish.

The evidence under Criterion C

Criterion C:

Evidence of dissatisfaction with the current grading standard from those who use the qualification, and those responsible for maintaining the standard

i. The views of those who use the qualification

Would users of the qualification support a change to grading standards?

The primary purpose of A level modern foreign language qualifications is to support entry to UK Higher Education institutions. For our previous investigation into grading standards, representatives from Higher Education were an obvious ‘user’ of these qualifications whose views we could seek. Given the broader range of purposes of GCSE qualifications, we recognised that the evidence under this criteria would need to reflect a broader range of views. We approached this question through the lens of the Common European Framework of Reference for Languages (CEFR). The purpose of doing so was to describe current performance and assessment standards at key GCSE grades using the CEFR ‘metalanguage’, which would allow comparison with equivalent qualifications taken elsewhere. This would also provide a meaningful context in which to discuss the suitability of the current grading standards in relation to performance standards in GCSE modern foreign languages (Curcin and Black 2019a).

This was accomplished using a study where performance and assessment standards at the key grades (grades 9, 7 and 4) in GCSE French, German and Spanish on the summer 2018 tests were notionally linked to the CEFR scale. Initially, content mapping (relating the construct and content of the GCSE to the CEFR) was carried out for each subject by a CEFR expert and a subject expert. Then panels of 13 experts (including CEFR, Higher Education and subject experts, A level teachers and exam board representatives) carried out several different benchmarking and linking activities within each subject.

For writing and speaking, experts rank ordered, in terms of overall quality, series of GCSE performances (at grades 9, 7 and 4) interspersed with performances previously independently benchmarked on the CEFR scale. This created an overall performance quality scale on which the relative position of the GCSE and CEFR performances was determined, and CEFR-related performance standards at grades 9, 7 and 4 were extrapolated from this. For reading and listening comprehension, the

experts conducted a 'standard linking' exercise to rate each mark point on the tests in terms of the CEFR levels. CEFR level cut scores were derived from these ratings and grades 9, 7 and 4 related to these in terms of proportions of marks on the test needed to achieve each. The linking results at component level were then averaged to get a qualification-level estimate of the mapping of each grade to the CEFR level.

The study found that performance standards between the three languages are reasonably aligned at qualification level despite some component-level inconsistencies. The results suggest that grade 4 is around high A1 level for Spanish and mid A1 level for German and French. Grade 7 is around mid A2 level and grade 9 around low B1 for all languages (see Figure 6 below). This result accords with the results of the content mapping, which suggested that each of the three GCSE MFL specifications assessed most of the skills up to A2+ (high A2) level, with some aspects of language competence assessed up to low B1 level.

While a degree of consistency across languages is perhaps to be expected given that these assessments are supposed to be developed based on specifications that are very similar in terms of content and implicit demand, there is no particular reason why we should expect the performance standards for different grades to be perfectly aligned across languages.

The results of this linking study essentially give an indication of where GCSE assessments are pitched and which performance standards are represented by different GCSE grades, using the language of the CEFR descriptors. However, this link is not a statement of what the GCSE standard should be, but an approximate description of what the performance and assessment/grading standard currently appears to be, using the language and descriptors of the CEFR.

To help determine whether the current standard was consistent with stakeholder expectations, exam board awarders (as custodians of the standard), as well as a wider stakeholder group (the majority of whom took part in the linking study panels) were asked to complete an online survey following the completion of the CEFR linking study. The survey aimed to establish what they thought the current performance standard is and what it should be for grades 7 and 4, using the metalanguage of the CEFR descriptors (Curcin and Black, 2019b).

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.	
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.	
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.	
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.	
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.	
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.	

Figure 6: CEFR benchmarking of GCSE French, German and Spanish grades 9, 7 and 4

The majority stakeholder view of the current performance standards at grades 7 (A2) and 4 (A1) broadly aligned with the levels identified in the CEFR linking study, although awarders, possibly due to little familiarity with using the CEFR scale, may have overestimated current candidate attainment in relation to the CEFR (A2 for grade 4 and B1 for grade 7). However, the majority of the stakeholder responses,

including those of awarders, also suggest that the current performance standards are lower than they wished they were. This is in the opposite direction to what we might have expected given all the stakeholder evidence considered under criterion D below, which suggests that French, German and Spanish should be made 'easier'.

It is difficult to readily reconcile how it might be possible to both increase performance standards as well as lowering grading standards. The majority of the stakeholder evidence and perceptions stem from the statistical outputs and interpretation of these. From this "statistical" point of view, the grading standards are perceived to be 'severe'. However, this survey focused on performance standards, arguably the key output of GCSE modern foreign language qualifications, (i.e., what students achieving different grades are able to do with language and what the GCSE should equip them to be able to do). This is perhaps unsurprising given respondents were given the opportunity to express a view on where they would wish standards to be.

Nonetheless, on the face of it, the perception that respondents would want to see higher performance standards, while the majority of evidence reviewed against criterion D suggests grading standards are perceived to be too high, appears contradictory. However, it is important to recognise that these views relate to stakeholders' aspirations. In addition, it would be possible to make changes to the qualification that would change what achieving a particular grade represents in terms of what students demonstrate in their assessments.

Such changes might be in the nature and focus of aspects of the assessment, and/or the focus of curriculum content so that there is greater focus in the qualification in developing those skills that stakeholders and subject experts consider to be the most important for students to develop. In such a situation, stakeholders might then be more content with what students at a particular grade achieve, potentially even in a context where grading standards had been relaxed slightly on a statistical basis, but such considerations are beyond the scope of this work.

We have raised the concerns that stakeholders who took part in the CEFR mapping held (and also those noted in relation to criterion D below), with the Department for Education, which is responsible for GCSE subject content.

ii. The views of the exam boards, and specifically the judgements of examiners responsible for making awarding decisions

Would exam boards and their awarders support a change to grading standards?

Two of the three exam boards offering GCSE modern foreign languages responded to our request for evidence. We asked exam boards to contribute their views on the acceptability of an adjustment to grading standards in these subjects. We also asked exam boards to gather evidence from their senior examiners responsible for awarding using a questionnaire which sought their view of the appropriateness of performance standards at judgmental grades 7 and 4 (those set by examiners with reference to archive scripts, rather than arithmetically calculated).

Awarders were also asked whether they felt the minimum performance required to achieve these grades had changed over the past ten years, if they felt an adjustment to the current standard would be acceptable, and whether they regarded the current standard as potentially having any impact on the wider education system. They were also asked their view of the likely outcome of any change to grading standards.

We expected the exam boards to take these views into account when submitting their organisational responses, as it is awarders who are the 'custodians' of the standard responsible for ensuring consistent performance standards are maintained year-on-year through reference to statistics based on prior attainment and archive scripts.

One exam board opposed any adjustment to grade boundaries on the basis that the current grades reflected the appropriate and expected performance standards for these qualifications. That board also counselled against any comparisons with performance standards in other, non-modern foreign language GCSE qualifications – making the point that GCSE French, German and Spanish require candidates to demonstrate specific skills which are not required in all subjects, and which are tested using different assessment methodologies. The board concluded that it is extremely challenging to compare the 'difficulty' of reaching a defined level of performance in different qualifications, and indeed it may be fundamentally inappropriate to do so.

The board also expressed the view that the perceived 'difficulty' of the subject should be considered in the context of changes to the subject content and assessment, which they felt were likely to have a negative impact on candidates' experiences of these qualifications. The board felt that these issues might be more productively addressed through a review of the requirements for the GCSEs.

The other exam board felt that there was strong evidence in favour of an adjustment against each of our criteria. It noted that there were possible counter-arguments in relation to some evidence, but on balance the board was satisfied that the strength of evidence was such that Ofqual should adjust grading standards in modern foreign languages in view of the strength of feeling amongst stakeholders and the risks to the subject posed by the long term decline in uptake. Amongst the evidence the board cited, the consistency of different statistical measures which suggest that these subjects are more severely graded than other GCSEs; the strong support from teachers and representatives from higher education for an adjustment; evidence of the importance placed on language skills by business and industry; views of its

awarders; and the fact that previous interventions to improve language uptake and performance through other approaches (such as the introduction of controlled assessment for Speaking and Writing as a result of recommendations from the 2007 'Languages Review') had not had the desired effect.

While the board acknowledged the possibility that the perception of severe grading of languages at GCSE could stem from other factors (for instance, motivation or teaching quality), it also noted the difficulty of untangling the various factors which could be giving rise to the perception of severe grading to an extent that a definitive decision as to its actual presence or absence could be reached. Despite citing potential negative impacts resulting from an unwarranted adjustment to standards, such as increased entry requirements at A level; complacency about addressing other issues impacting on language uptake; and the potential for setting a precedent that declining subject popularity or issues with teaching and learning should be managed through changes to grading standards. On balance, the board viewed the situation in GCSE modern foreign languages as being suitably distinct (and the potential consequences for a subject which government and wider society regard as important, sufficiently dramatic) that an adjustment would be appropriate regardless of the risk that severe grading – perceived or otherwise – may not be the cause.

The board did not offer a view on the scope of adjustment which might be made or how this might differ between subjects. It did, however, recommend that if standards were to be aligned that this should be accomplished through an incremental adjustment over a number of years to bring grading standards statistically into line with those in geography and history, rather than attempting to make a one-off adjustment which would unduly benefit one year's cohort in comparison to the previous year's.

A summary of the views of awarders from each board is presented below by subject.

French

Awarders' views differed between exam boards, with those at one board generally regarding the standard as appropriate and awarders at the other generally of the view the standard was too severe. At neither board were awarders' views unanimous however.

Awarders from the exam board where it was generally felt that standards were currently acceptable typically felt that the standard at grade 7 was generally appropriate, also pointing to script evidence that standards were comparable to previous years. One commented that the standard reflected what was appropriate for students who had studied the language for five years. At grade 4, the majority of awarders felt that the standard was appropriate, but others argued either that it was too severe or too lenient.

Awarders at this exam board did not agree whether the minimum performance standards at grades 7/A and 4/C had increased, decreased or stayed the same – but

those who considered standards had changed generally attributed this to changes to the subject content rather than any clear divergence in performance standards. Similarly, awarders were divided over whether an adjustment to grading standards would be acceptable. Some argued that it would, despite the current standard being appropriate, if it were to address misalignment with other subjects. Others felt that it would have a negative impact on students' preparedness for A level, or lower grade boundaries resulting in a loss of discrimination.

The majority of awarders considered standards set in the subject were having an impact on wider society, but generally due to 'perceived' rather than actual 'difficulty'. Most awarders felt that an adjustment to standards would be likely to have a mixed impact – citing concerns about preparedness for A level even if it did lead to an increase in entries. One awarder felt it would be positive if it improved the perception of the subject in the eyes of school senior leaders, but another was concerned about possible confusion arising from a lack of comparability with standards set in previous years. One awarder worried that adjusting standards would mean losing improvements made in overall student performance made since the qualifications were reformed.

Awarders from the board that generally felt that the standard was too demanding considered this was the case at grade 7 and grade 4. They argued variously that changes in subject content had made the standard harder to achieve for students, or that languages were intrinsically harder than other GCSE subjects and that this should be taken into account at awarding.

One examiner argued that the standard of performance at the grade 7 and grade 4 boundaries seen at awarding was sufficient to meet some of the grade descriptors for grades 8 and 5 respectively, but that they thought they were prevented from recognising this due to an inability to deviate from the statistically recommended boundaries. (In such a scenario we would expect the board in question to make a case for an out of tolerance award).

Similarly, when asked to consider whether standards had changed over the past ten years, all awarders at this exam board felt that the minimum performance standard required for both grades 7/A and 4/C had increased. While for most this was again the result of changes to subject content and assessment format, one awarder argued that in their view minimum performance standards in languages had always been more challenging than other GCSEs. Another felt that the standard required was higher when compared with that ten years ago.

Given this, the majority of awarders from this exam board agreed that an adjustment was required to grades 7 and 4 – arguing variously that this was necessary to reflect the changes to subject content, was required to bring the subject into line with other GCSEs, and in one case that if it were sufficiently small it would not be noticeable in terms of performance standards. All but one felt that the current standard was having a detrimental impact on the subject, while the dissenting awarder felt that an adjustment was desirable but unlikely to reverse the decline in entries.

The same proportion of awarders argued there would likely be a positive washback effect from an adjustment to standards in terms of addressing “historic and obvious severe grading” which made it harder to achieve high grades. One awarder suggested that while this would create an incentive for schools to study the subject there would be a risk that other stakeholders may regard the subject as consequently being too easy, which would be detrimental to the status of French.

German

Exam board awarders were strongly polarised on the acceptability of an adjustment to standards at grades 7 and 4 in GCSE German. Awarders from one board universally felt that the standards at both grades were appropriate, and in particular that the current standard at grade 7 (as a judgmental grade from which arithmetic boundaries are calculated) was necessary to ensure that students receiving this grade were adequately prepared for the demands of A level.

While one awarder felt that there was perhaps an argument that the minimum performance standard at grades 7/A and 4/C had increased compared to ten years ago, this was directly attributed to the changes to the content demand of the qualification as a result of GCSE reform. Others disagreed. None of the awarders felt that an adjustment to standards would be acceptable, unless it was small enough to fall within the tolerances usually allowed to the exam boards at awarding.

The majority of awarders agreed that the German GCSE was ‘difficult’ for students, but generally attributed this to the removal of alternative language qualifications for less academic students, requiring them to sit the GCSE instead, rather than the GCSE standard itself. One disagreed, arguing that perceived difficulty was far more likely to be the result of approaches to teaching and limited exposure to languages outside of schools.

There was no clear consensus that an adjustment to standards would help to increase uptake. Some awarders were concerned about the impact an adjustment would have on the perception of the subject and the precedent it would set.

Awarders from another exam board, on the other hand, were generally strongly supportive of an adjustment to standards – although this was not a unanimous view. The majority felt that the standard set at grades 7 and 4 was now too demanding, with only one awarder arguing that these standards were appropriate. The inappropriate level of demand was felt to be more of an issue at grade 7 than at grade 4, and was attributed either to historic misalignment of standards or changes to subject content as a result of the reforms. However, one awarder argued that the subject was intrinsically more challenging than other GCSEs and even than other languages – suggesting that aligning to other GCSEs based on relative ‘difficulty’ would inevitably mean candidates would demonstrate a lower standard of performance in German than in either French or Spanish.

Similarly, most awarders felt that the minimum performance standard required for grades 7/A and 4/C had increased (although again by less at grade 4 than 7). While some attributed this to the implications of structural changes to the qualification, others argued that the standard required now was notably harder when compared to 2007/2008 scripts.⁷ One awarder disagreed, feeling strongly that the standard had been maintained, and that it would not be appropriate to adjust standards at grades 7 or 4. The other awarders argued students taking GCSE German were not competing on a level playing field with those in other subjects, and felt that an adjustment would be likely to go some way to helping reverse the decline in entries.

Awarders were also split on the same lines on whether the standard set currently in the qualifications was having any wider impact. One awarder argued that it was not, and the perception that German was a 'hard' GCSE was based on factors other than grading. The remaining awarders felt that the grading standard was leading to a number of negative impacts, including declining entries overall and restrictions schools placed on the type of student they allowed to study the subject.

Views on the potential washback effect of an adjustment to standards were more mixed however. While the majority still felt that an adjustment would have a positive impact if it served to increase entries and perceived it as being fairer to students, one concluded that it was not likely to lead to any significant increase in entries and another felt that it would make progression to A level more challenging.

Spanish

There were similarly mixed views on the acceptability of an adjustment to grading standards in Spanish at one board. While awarders were unanimously of the view that the grading standard at grades 7 and 4 was appropriate, awarders differed on whether they felt that the minimum performance standard for grade 7/A had increased, decreased, or stayed the same over the past decade. The majority of awarders argued the latter, although one was of the opinion that the standard was demonstrably higher than ten years ago based on comparison of archive scripts.

At grade 4/C awarders were more generally in agreement that the standard had remained unchanged, although one awarder felt that the changes to the qualification at the time of reform meant that it was now particularly demanding for foundation tier candidates. Awarders divided evenly over whether an adjustment would be acceptable. Where awarders argued for an adjustment, the view was expressed that this should be at grade 4 only and not at grade 7. One awarder supported an adjustment but was concerned about the impact that this would have on comparability with qualifications awarded in Wales. The remaining awarders felt that the current standard was correct and grade boundaries should not be lowered.

⁷ This contrasts with the arguments of some stakeholders that longstanding 'severe grading' has been inherent in these qualifications since their inception, and potentially carried forward from O level.

Most awarders felt that the ‘difficulty’ (perceived or otherwise) of GCSE Spanish was having an impact on entries, but that this was either a misconception, or the result of interactions with accountability measures and not that the standard itself was inappropriate for a GCSE. One awarder felt that issues in relation to language had no relation to standards, and were instead the result of a range of factors which had damaged the status of the subject (including its optional status).

There was also no clear consensus on whether an adjustment to standards would be likely to have any wash-back effect – some suggesting that it could lead to an increase in uptake, while others felt that it would be unlikely to have any impact or could damage the perception of Spanish if it was regarded as a “dumbing down”.

Awarders at the other exam board were unanimously in favour of an adjustment to grading standards. All felt that the standards set in the subject were too demanding at both grades 7 and 4, although this was attributed by some to the fact that the subject is intrinsically more demanding than other GCSEs, or to the implications of changes to the subject content. A significant proportion, however, felt that there was evidence that candidates on the boundaries for grade 7 and 4 were in fact displaying levels of attainment matching the descriptors for grades 8 and 5.

This board’s awarders were also universally of the view that the minimum performance standards for grades 7/A and 4/C had increased within the last ten years. While this was generally attributed to the greater demands of the subject content and changes to assessments some argued that there was clear evidence that standards at these grades were higher than those demonstrated ten years ago. As noted above, the exam board in question could have made a case for alternative grade boundaries in this subject from those suggested by the statistics if it had wished.

There was uniform agreement amongst awarders at the second board that an adjustment would be acceptable. In contrast to awarders’ views at the first board, they felt that there was greater need for alignment at grade 7 as opposed to grade 4. Awarders across the board also felt that the current standard was having a negative impact on the subject, citing issues with teacher recruitment and retention issues, declining uptake due to perceived severe grading, a decline in dual linguists at A level, and the number of schools offering no modern foreign language GCSEs.

Generally awarders felt an adjustment would be beneficial in terms of addressing a perceived historic injustice and potential increased uptake at A level and university, although one awarder was concerned that while an adjustment might lead to more students taking the subject at GCSE it might also result in students being less prepared at A level.

Conclusion

The CEFR study indicated that qualification standards in GCSE French, German and Spanish were generally well aligned with each other (and in fact, potentially a little

more demanding at grade 4 for Spanish than either French or German). To our knowledge this is the first time that such a benchmarking activity has been undertaken for GCSE, and there is no defined expectation of CEFR equivalence in the GCSE modern foreign language subject content. In the survey, the majority stakeholder view of the current performance standards at grades 7 (A2) and 4 (A1) broadly aligned with the levels identified in the CEFR benchmarking study, although awarders, possibly due to little familiarity with using the CEFR scale, appeared to overestimate current candidate attainment in relation to the CEFR. However, the majority of the stakeholder responses, including those of awarders, also suggest that the current performance standards are lower than they thought they might ideally want them to be when considered against the CEFR scale. This is perhaps unsurprising given respondents were given the opportunity to express a view on where they would wish standards to be.

Nonetheless, on the face of it, this perception that respondents would want to see higher performance standards, while the majority of evidence reviewed against criterion D suggests grading standards are perceived to be too high, appears contradictory. However, it is important to recognise that these views relate to stakeholders' aspirations. In addition, it should be noted that the majority of the stakeholder evidence and perceptions about severity of grading standards stem from the statistical outputs and interpretation of these. However, this survey focused on performance standards, what students achieving different grades are able to do with language and what the GCSE recognises them as being able to do. It may also be that this tension reflects stakeholder preference to see some change to the focus of aspects of the assessment, and/or the focus of curriculum content, so that there is greater focus in the qualification in developing those skills that stakeholders and subject experts consider to be the most important for students to develop. We have passed this evidence to the Department for Education to consider.

The CEFR work shows that students are producing work which is of sufficient quality to register against the framework at grades 9, 7 and 4. This is significant in relation to our duty to have regard to international comparability of qualifications standards. The study showed that the current performance standard at the lowest judgemental grade in each subject (grade 4) still registers comfortably on the CEFR scale, suggesting against this criterion that there may be scope for a cautious relaxation of grading standards without changing the broad CEFR levels which describe performance at key GCSE grades.

Our decision in relation to this criterion is also based on consideration of a wider evidence base. We also sought the views of exam boards, and asked their awarders on the acceptability of adjusting grading standards through a separate questionnaire. Two exam boards gave us their views. One felt the current grading standards are appropriate. The other argued they should be lowered as other attempts to address perceived severe grading through changes in qualification design and attempts to improve students' experience of taking these qualifications had failed to increase

entries. Their awarders also disagreed along similar lines. Some of the awarders thought it would be acceptable to adjust grading standards in German; more, although not all, thought the same for French and Spanish, although they did not agree on the grades that should be adjusted. However, most awarders felt that relaxing standards might help increase uptake; even those who considered the current grading standards to be correct.

While the evidence under this criterion was mixed, we judge that there is sufficient weak evidence in support of a potential adjustment to grading standards in all three subjects to conclude that users and awarders of these qualifications would be likely to accept a relaxation of standards at key grades.

The evidence under Criterion D

Criterion D:

The likely benefit to users of the qualification and society as a whole from a change to grading standards must outweigh any potential negative effects

i. There is evidence of support from users of the qualification for any change

Is there evidence of strong support for an adjustment to grading standards?

Teachers' perceptions of severe grading in GCSE French, German and Spanish, and the claims that this is related to the current declining entry at secondary and university level are widely held and long established. These claims are acknowledged in, but predate, the 2008 QCA report *Grade standards in GCSE modern foreign languages* and the 2007 Languages Review by Lord Dearing .

In 2016, as part of [our initial policy consideration of the challenges presented by inter-subject comparability](#), we conducted a [survey](#) to gauge public feeling about a number of different potential policy approaches to addressing inter-subject comparability. These were discussed at our February 2016 conference. This survey received 216 responses, with the majority of respondents (42%) favouring greater alignment of subject 'difficulty' being achieved through adjustments to grading. A significant number of responses to this survey were from teachers or those with an interest in modern foreign language learning, concerned about the perception of severe grading.

In 2017, the Association of School and College Leaders (ASCL) surveyed members of subject associations and other stakeholder organisations regarding GCSE modern foreign languages. Of the approximately 2,600 respondents:

- 96% believed that these subjects were severely graded
- 98% were in favour of action by Ofqual to align grading standards in EBacc subjects so that there was no systematic variation in outcomes
- 81% argued that it had resulted in a decline in the numbers studying the subject at Key Stage 4

- 96% felt that it had a negative impact on students' perceptions of the subject and progression to A level or Higher Education

To help establish whether there is a compelling case to adjust grading standards in these subjects, we wrote to stakeholders in March 2019 asking for their views on the evidence to include when considering the case for an adjustment. In response, a range of stakeholders (from Higher Education, a learned society and ASCL itself) quoted ASCL's 2017 survey as evidence supporting the need for a change to grading standards. One such respondent felt the "negative impact of maintaining the current severe grading situation on the confidence of thousands of pupils, parents, teachers and senior leaders in the GCSE modern foreign language qualification is compelling evidence" that qualification users would welcome an adjustment.

The British Council's *Language Trends 2019* survey report found that 62% of independent school teachers surveyed and 59% of state school teachers cited exam marking and grading as a significant issue. Comments noted "the difficulty of achieving a high grade in a language compared to other subjects, and the perception that students are not making good progress" as being a major area of concern. Respondents, including representatives from the Higher Education sector, and a British learned society, cite this survey report as strong evidence of widespread concern about grading standards in GCSE modern foreign languages.

In May 2019, 152 academics from 36 universities wrote an open letter to *The Guardian* asserting that GCSE and A level modern foreign language exams were graded too severely, and asserting that this is the cause of the decline in uptake of languages. These university teachers described [modern foreign language] GCSEs as "vital to the pipeline" for A levels, degrees and teacher training, and called for an urgent adjustment to grade boundaries for modern foreign language exams at both GCSE and A level.

A significant majority of the detailed responses we received from stakeholders from the Higher Education sector reflect the academics' view in the open letter that there is a need to adjust grading standards. Many raised concerns about what they felt to be severe grading in GCSE modern foreign languages and called for a one-off adjustment to grade boundaries. The reasons cited included "strong, widespread support" from the higher education community and "overwhelming support" from the language teaching community for such a change. They often referenced ASCL as actively calling for an adjustment on behalf of head teachers, and described the "negative impact of maintaining the current severe grading situation on the confidence of thousands of pupils, parents, teachers and senior leaders in the GCSE modern foreign language qualification" as "compelling evidence" that qualification users would welcome an adjustment. They offered, as the primary reason to adjust grading standards, the belief that "it is the right and fair thing to do" and a view that pupils should not "be judged more harshly in languages than in other subjects".

The view that severe grading is the cause of a decline in uptake of GCSE modern foreign languages (expressed in the open letter to *The Guardian* quoted above) was echoed by the majority of other stakeholders who provided evidence, including representatives from schools, Higher Education and headteachers' associations. Other concerns stakeholders felt to be associated with severe grading, and/or perceptions of severe grading and declining GCSE modern foreign language entries, included falling teacher recruitment and teacher retention, with a consequent effect on the pool of language teachers in further and Higher Education. One organisation, representing the Higher Education sector, described severe grading as the major barrier to the uptake of languages and linked it to a loss of public confidence in the grades awarded in languages which it believes has reached parents and pupils.

In addition, a French cultural institution argued that its analysis of GCSE results statistics has found "a systematically higher proportion of students getting top grades... in GCSE Spanish than in GCSE French or German". It felt this imbalance disadvantages students taking French and German and may deter GCSE uptake.

All individual respondent schools raised concerns (to a greater or lesser degree) about severe grading in GCSE modern foreign languages and/or the need to adjust grade boundaries. A school with a strong focus on languages, which describes itself as performing significantly above the national standard and previously achieving "impressive [modern foreign language] results", provided evidence of its pupils' "underperformance in [modern foreign languages]" compared with their average points score across all other subjects in 2018. The school regards this evidence as being of a suitably large data set not to be considered anecdotal. It believes there is a compelling case that modern foreign language GCSE grading standards are harsh and need adjustment to make them comparable to other GCSE subjects. Another school felt GCSE grade boundaries for the different grades to be far too wide.

The imbalance in modern foreign language results highlighted by the French cultural institution (referred to above) was also suggested by a school which argued that French is graded more severely than Spanish. This school suggests we are allowing French to become obsolete by demeaning its value and calls for parity between the languages. Another school found that, compared to French and Spanish, its students' German results fared poorly. It perceived a "reluctance to allow students to achieve well in German".

Both German language subject association respondents raised concerns about students receiving lower grades in GCSE German than in other subjects. They shared perceptions that languages are difficult, and felt this affected what schools wanted to offer. They expressed belief that "fairer grades at GCSE" will increase A level uptake and halt the decline of German.

Views from headteachers' associations were divided, with two calling strongly for an adjustment to grading standards, another warning against changing the standard, and one suggesting that no single piece of evidence could demonstrate the case for an adjustment and cautioning that the evidence needs to be carefully weighed up to

decide whether change is needed, with statistical evidence receiving the highest weighting.

One headteachers' association provided what it believes to be clear evidence from its schools of students having achieved lower grades overall in modern foreign language GCSEs compared to other subjects for many years. It suggests this disparity became "significantly greater" after the introduction of reformed GCSEs. It argued this data is compelling evidence that French, German and Spanish are more severely graded than other subjects, showing a clear need to adjust grading standards. The data came from schools where, it asserts, languages provision is strong with excellent, very experienced teachers who were well placed to respond to the new specifications and requirements of the reformed GCSEs. It cited the decline in students studying modern foreign languages at GCSE and beyond as a major national concern, and argues that adjusting grading standards is crucial to addressing harsh grading and reversing this decline.

Not all stakeholders supported a change to grading standards. Some expressed no clear view. Others warned against a change.

For example, in contrast to the views noted above, one headteachers' association that warned against changing the standard felt there were multiple reasons not to make an adjustment. These included the need to maintain a degree of rigour in the qualifications; the challenges of conceptualising what is meant by "difficulty" in comparing different subjects and how best to measure this; variance of relative difficulty at different grades within a subject; and a danger that changing the grading standard to make a subject easier may encourage students to pursue a subject they then struggle to study at a higher level. This association suggested that the fall in entries for GCSE French and German could be due to how these subjects are taught, or their reducing relevance to a world looking towards other countries as emerging cultural and economic powers, and that changing the grading standard will not reverse these trends.

In addition, one teachers' union expressed caution about whether or not issues with a qualification's difficulty level should be addressed through its design, content and assessment arrangements and expressed concerns about the extent to which an adjustment to grading standards might be perceived as an inappropriate response to behaviours which may ultimately be driven by government performance measures. They felt this could undermine confidence in Ofqual's role as an independent regulator. However, the stakeholder argued this should be balanced against the threat to public confidence in modern foreign language GCSEs, and broader comparability of qualifications standards, arising from the perception that these subjects are severely graded.

Overall, there was strong stakeholder support, particularly from the modern languages community, for an adjustment to grading standards, especially in French and German. In particular, many stakeholders who did support a change expressed strong views in favour of adjusting grading standards to bring modern foreign

languages in line with other subjects and ensure a “level playing field”. However, a minority of respondents, mainly non-modern foreign language academics or stakeholders, expressed contrary views or suggested the need for caution. Some respondents suggested any perceived ‘difficulty’ of these subjects could be a result of aspects of these qualifications other than grading. Others questioned the likely value of an adjustment, and raised concerns about the potential for unintended negative consequences.

Responses also show that while language stakeholders broadly agree that “modern foreign languages” (taken collectively) are severely graded in their view, perceptions of the relative difficulty of GCSE French, German and Spanish sometimes vary. There was a view, however, that Spanish is perceived to be less severely graded than either of the other two languages by some stakeholders.

Stakeholders arguments tended to be focused on grading standards at grades 4/C and above, with few concerns and little evidence of issues with grading standards at the lower grades presented.

ii. There is no reason to believe that there would be a detrimental impact on the extent to which the subject fulfils the defined purposes of the qualification

What evidence is there that an adjustment to grading standards would be detrimental to the defined purposes of GCSE qualifications?

The defined purposes of GCSE qualifications are as follows:

- To provide evidence of students’ achievements against demanding and fulfilling content
- To provide a strong foundation for further academic and vocational study and for employment
- To provide (if required) a basis for schools and colleges to be held accountable for the performance of all of their students

There is a risk that adjusting grade standards could potentially undermine the technical functioning of the assessments, and the performance standards different grades represent. This suggests a need to proceed with caution in terms of any adjustment, to ensure that this does not happen to an unacceptable degree.

Some awarders expressed concerns that an adjustment to standards would result in GCSE students being less well prepared for progression to A level and Higher Education, and could impact negatively on retention rates. This was a minority view however, and was not consistently expressed amongst awarders at both exam boards who participated in the survey or in all the languages. This was generally felt to be more of a concern in GCSE French and German than in Spanish.

One exam board expressed a similar view, arguing that while addressing perceived severe grading through an adjustment to standards might lead to a temporary increase in numbers at GCSE in the short term, it could impact negatively on progression and retention at A level and in higher education if students completing the GCSE were less prepared or higher levels of study. It was argued that this could eventually lead universities to increase their entry requirements for modern foreign language courses. The findings from our previous research into [HE perceptions of A level grading standards](#) suggest this is unlikely.

We heard from stakeholders that the interaction of potential grading severity with government performance measures in the form of Progress 8 and EBacc is seen to penalise schools who offer languages at GCSE, acting as a disincentive to offering them. Some expressed concerns that their analysis of value added data indicates that students systematically perform worse in modern foreign languages when compared to other optional subjects counting within the EBacc. This analysis presumes that students of a given level of prior attainment should make similar levels of progress in all subjects. One of the stated purposes of GCSE qualifications is to provide a basis for schools and colleges to be held accountable for student performance. As those accountability measures do not make a distinction between the same grade awarded in different subjects, beyond expecting achievement in particular subject groupings, it might be seen as problematic for such measures that GCSEs in modern foreign languages are not aligned on a statistical basis with other subjects that count in the measure. An adjustment to grading standards could address this – but would mean Ofqual potentially prioritising the accountability purpose over the other stated purposes of GCSEs. On the other hand, if this argument is accepted, the government could adjust accountability measures to reflect the differences in average student progress in different GCSE subjects.

iii. There is no reason to believe any change would have a detrimental impact on performance standards, for example by decreasing the level of cognitive demand in comparison to other cognate subjects

What evidence is there that a change in standards would be likely to have a detrimental impact on performance standards?

Given the significant limitations of statistical measures of subject difficulty, we judge that it would be inappropriate to base any adjustment solely on the basis of such measures. Some of the measures suggest adjustments which are more significant than others, so in any case we need to exercise judgement as to how they are used. As statistical difficulty is wholly relative and does not reflect actual performance standards within a subject, any adjustment made on a statistical basis alone inevitably risks an impact on performance standards if not made with reference to expert judgement.

For instance, statistical modelling, set out in our statistical evidence report (He & Black (2019)), based on 2016 NPD data of the grade boundary adjustment necessary to bring GCSE French, German and Spanish into alignment with the statistical average for all subjects, suggests that the scale of change would be significant at certain grades. This would not necessarily align with other evidence reviewed. It may also be potentially unacceptable to some stakeholders, as indicated by the caution counselled in some of the responses we received to our request for views; although the stakeholder evidence does suggest a significant number would be likely to accept such changes. It would also entail increasing grade boundaries at some grades in Spanish (raising performance standards) which, given the general perception of modern foreign languages as severely graded, would concern some stakeholders.

The findings from the CEFR survey of awarders indicated that they deemed the actual and aspirational standard of candidate performance to be higher than their actual performance, as indicated by the research. This may counsel against an adjustment to standards as it suggests that students are already failing to reach the level of competency expected by awarders when measured against the CEFR. The limitations of this research are considered in greater detail in the accompanying report. As discussed earlier, there are currently no stipulated or commonly accepted expectations of CEFR equivalency amongst GCSE stakeholders at either qualification level or at specific grades in either the legacy or reformed GCSEs. In an attempt to address this, we broadened the CEFR survey to include other stakeholders, including representatives from the exam boards, higher education, language institutes, subject associations, subject experts and teachers.

The results of the survey indicated that the majority of the respondents perceived current performance standards at grades 7 and 4 to be lower than they should be – that is, that their expectations of the performance grade 7 or grade 4 candidates should demonstrate performance as described by a higher grade on the CEFR. This conflicts with the expectations set by stakeholders that languages should be made ‘easier’. It is difficult to readily reconcile how it might be possible to both increase performance standards as well as lowering grading standards. Note however that the findings of the survey report concluded that this might be reflective of a desire for

changes in the nature and focus of aspects of the assessment, and/or the focus of curriculum content. This aligns with the results of the content mapping and some of the qualitative findings of the CEFR linking study, which suggested that some aspects of current assessment practices and curriculum content may be adversely impacting on student performance when considered against those elements prioritised in CEFR.

This survey also suggests that there are a range of different perceptions of performance standards amongst different stakeholders. Overall this is a relatively weak source of evidence given the limited number of stakeholders who took part, and one we have considered alongside the greater range of stakeholder views summarised above. The range of views expressed (and the apparent contradictions in some of the findings of the various CEFR work) again suggests that we should be cautious in determining the scale of any adjustment we may decide to make.

iv. There is no reason to believe that there would be a significant detrimental impact to other parts of the education system as a result of an adjustment

What evidence is there that an adjustment to grading standards might have a significant detrimental impact to other parts of the education system?

We had little evidence to consider under this sub-section of the criterion. Nonetheless, we did engage directly with a broad range of stakeholders in the course of this work. While many stakeholders, especially those in the modern foreign languages community, stressed their view that adjusting grading standards would address a historical inconsistency and a source of disadvantage in relation to other GCSE subjects (and could not see a potential negative impact in doing so) we also heard from organisations who were sceptical about the advisability of an adjustment. Some noted potential risks and urged us to be cautious in relation to adjusting standards, as noted above.

While we recognise the need to be particularly mindful of the risk of a negative impact on overall public confidence in qualifications in any adjustment we may make, particularly if any adjustment were seen to be disproportionate, we did not identify any other significant additional risks in this area.

Conclusion

We judge that the evidence generally supports an adjustment to grading standards in all three subjects. Stakeholders' submissions, alongside other sources of evidence and the Higher Education perceptions study we conducted as part of our A level work, suggest that an adjustment to standards would be broadly accepted by most

users of the qualifications and would be unlikely to lead to any counter-productive response from Higher Education. Indeed, it would likely be widely supported. There were also some who felt that Spanish was less severely graded than French or German.

While any adjustment to grading standards will also inevitably have some impact on the overall level of performance demonstrated by students obtaining each grade, we are satisfied that it would be unlikely to be detrimental to the defined purposes of GCSEs as long as the adjustment is appropriately calibrated. It may even make these qualifications more effective in providing a basis for school and college accountability, as it will move modern foreign languages closer to the statistical mean. However, there were stakeholders who expressed potential concern about the negative impact that such an adjustment could have on other aspects of the system, which therefore suggested we should exercise some caution regarding the nature and scale of any adjustment.

Statistical modelling of the impact of aligning standards on grade boundaries suggested that there could be a notable drop in performance at certain grades if we aligned to a statistical mean of all GCSE subjects. This would be much more pronounced in French and German than in Spanish. This would have an impact on the performance standard and therefore the interpretation of GCSE grades, in a context in which the evidence reviewed in relation to criterion C suggests awarders may already be overestimating levels of students' performance, and a very significant adjustment could lead to altering the relationship with the CEFR we have observed. A change in this resulting from lower performance would not be supported by our survey evidence about stakeholders' expectations. Some stakeholders also noted potential negative impacts of an adjustment on other aspects of the education system, although these were weighed against the potential threat to public confidence of the perception that French, German and Spanish are more harshly graded than other GCSEs.

Overall we concluded that this criterion was met for all languages so long as any adjustment was appropriately calibrated.

Decisions

Based on our consideration of the evidence we have assembled, judged according to our criteria, we have decided:

- The evidence in GCSE French and German presents a sufficiently compelling case to make an adjustment to grading standards

- In GCSE Spanish, the evidence is less strong and overall does not present a sufficiently compelling case. Therefore we will not make an adjustment to grading standards in this subject
- We will align grading standards in French and German with those in Spanish at grades 4 and above. We will not seek to make changes to grading standards in these subjects at grade 1, but changes made at grade 4 could have a small impact on grades 2 and 3
- We will determine with exam boards the technical detail of how this adjustment might be most appropriately implemented from summer 2020

Judgement about an adjustment to grading standards in French, German and Spanish

In coming to this decision we have balanced a number of different sources of evidence, taking into account the strengths and weaknesses of each. We are satisfied that a sufficiently strong case exists for us to adjust grading standards in GCSE French and German, as we set out in detail above in relation to each criterion.

In relation to criterion A, both of these subjects appear to be consistently harder than other GCSE subjects under statistical measures.

In relation to criterion B, both have also experienced a prolonged and significant decline in entries since a change in policy in 2003 meant that they were no longer compulsory at GCSE. Evidence suggests that teacher recruitment is becoming more difficult generally in these subjects (despite the relative qualification level of teachers appearing to improve in recent years).

For criterion C, while the evidence under this criterion was mixed, overall we are of the view that there is sufficient weak evidence in support of a potential adjustment to grading standards in all three subjects to conclude that users and awarders of these qualifications would be likely to accept an adjustment to standards at higher grades.

In relation to criterion D, there is strong stakeholder support from subject associations, some teacher organisations and Higher Education modern foreign language academics for adjusting standards.

The evidence is less strong in GCSE Spanish. There is stakeholder and awarder support for an adjustment to standards under criteria C and D. There is also some evidence to suggest that potential issues with teacher recruitment or quality may in fact be more severe in this subject than other languages under criterion B. However, in relation to criterion A, Spanish does not appear, statistically, to be consistently more severely graded than other GCSEs overall at the full range of grades and, in relation to criterion B, entries are not declining. Therefore we have concluded that there is not a sufficiently compelling case to adjust grading standards in GCSE Spanish.

Our statutory objectives include a duty to maintain public confidence in regulated qualifications, and to ensure they represent a consistent level of attainment between comparable qualifications. Having found that those criteria are fulfilled in the case of GCSE French and German, we regard the arguments to adjust standards in these subjects to promote public confidence and better align them with other similar qualifications, in line with our statutory objectives, to be sufficiently strong to justify doing so over the part of our objective which requires us to seek to maintain standards over time.

Judgement about the adjustment to be made in French and German

The ‘fuzziness’ of the statistical evidence, where different measures of relative subject difficulty indicate that French and German appear to be more severely graded than other GCSEs, but vary by how much and at which grades depending on the model used, makes it challenging to make fine judgments about exactly where the new grading standard should be set.

In determining an appropriate adjustment we are mindful that any adjustment should not assign arguably spurious precision to statistical measures of subject difficulty by attempting to align to a potentially volatile ‘mean difficulty’ and that such adjustments may not necessarily align with other evidence we have considered. We are therefore not confident that any of the approaches considered in our statistical evidence report (He & Black (2019)) would provide a sufficiently robust basis for an adjustment to standards.

In light of these concerns about the statistical evidence, we have resolved not to rely solely on statistical measures to determine the nature of the adjustment required in French and German. Instead, we have determined that it would be more appropriate to align grading standards in French and German to those in Spanish, for which GCSE and A level entries are steadily increasing. Doing so would allow us to accomplish what we judge to be an appropriate easing of performance standards.

Therefore we have concluded that there is a logical basis for adjusting grading standards in French and German so they align with those in Spanish. We note stakeholder arguments that unduly severe grading is a major factor in the decline in entries for modern foreign language GCSEs. Entries are declining in French and German, but not in Spanish. Statistical and stakeholder evidence also suggests that French and German are more severely graded than Spanish. According to this hypothesis, the difference between the apparent statistical difficulty of French and German on the one hand and Spanish on the other could be a factor in dissuading students from studying French and German. Aligning standards in French and German to Spanish would eliminate the potential impact of the difference in statistical difficulty upon entry.

In our view this is as far as the evidence supports the hypothesis of a negative impact on entries for modern foreign languages resulting from unduly severe grading, and therefore represents the furthest extent to which we can justify an

adjustment because we do not see an impact in the less severely graded Spanish. It also reflects that stakeholder concerns about severe grading were greatest for French and German. Some of the evidence under criteria B and D also suggests we should be cautious in terms of the scale of any adjustment we were to implement.

As we have no statistical evidence and few specific stakeholder concerns about the relative difficulty of grades below grade 4, we have determined there is not a case to adjust grade 1 in any of the three subjects, but changes made at grade 4 could have a small impact on grades 2 and 3.

Aligning outcomes in grades 4 and above in GCSE French and German to those in Spanish would make French and German less severely graded in statistical terms at all grades above grade 4 without, in our judgement, being likely to significantly change the minimum performance standards established in these subjects. We think it is also more justifiable than attempting to bring the standards closer to the mean of all subjects, given the volatility of that measure. These subjects share common government subject content requirements at GCSE, so there is greater justification for aligning grading standards in French and German to Spanish than there would be to another, non-cognate subject such as history or geography. We might also expect that students taking French or German will make similar progress in their studies to students studying Spanish, which might not be the case for other, less similar, subjects.

In considering the nature of the adjustment that we will expect exam boards to make in French and German, we have been mindful of the impact of our decision on students taking the qualification in different years.

In line with the approach we have taken in similar circumstances on previous occasions, for example when we announced an adjustment to grade standards as a result of the impact of native speakers in [A level French, German and Spanish in 2017](#), we will not require exam boards to retrospectively regrade students that took these subjects in earlier years. We do not normally consider it appropriate to require adjustments to historic awards. In any case, the evidence available would not allow us to identify for which years any retrospective adjustment should be made.

Instead we will be transparent about when adjustments to grading standards have been made, and that this will be for qualifications awarded from 2020. This will mean that those that use the grades, for example for selecting students for courses, can take this into account when considering students' results. In addition, depending on the exact size of the adjustment we determine is appropriate, we may require exam boards to introduce it incrementally over more than one year in order to make sure that the difference in standard between any two years is not too large. This will avoid undue unfairness between students in adjacent years of entry.

We have considered the arguments for also adjusting grading standards in other modern foreign language GCSEs, including lesser-taught languages such as Russian, Italian, Bengali and Persian. These share the same subject content (contextualised for each language) as French, German and Spanish. However the statistical evidence does not suggest that these other languages are severely graded. Although we know that the statistical evidence has its limitations, as set out in detail above, on the face of it there is no case to begin a detailed investigation into these subjects as our criterion A would not be met.

We explain how we will set about implementing our decision below.

Impact of an adjustment

We are cautious in predicting what the impact of the adjustment to grading standards is likely to have on the uptake of GCSEs in French and German. The evidence we have considered suggests the potential causes for declining entry (for both GCSEs and A levels) are numerous. Reasons are likely to include cultural attitudes to language learning, constraints on students' choices and the impact of government policies as well as perceptions of subject difficulty. Adjusting grading standards will, of course, address just one of these factors. We will adjust grading standards as far as is justified by the evidence and is compatible with our statutory objectives. We will not relax grading standards further if, regardless of the adjustment, entries continue to decline.

The majority of the statistical models of subject difficulty we considered in our report are relative measures. This means that GCSE French and German could continue to appear relatively severely graded in the future, or even become more so, for reasons unrelated to the actual grading standard applied, such as changing entry patterns in other subjects. Were this to happen in the future, on its own we would not consider it grounds to adjust grading standards further.

We believe aspects of the qualifications, beyond their grading standards, could contribute to perceptions that these subjects are 'difficult'. We have shared feedback we have heard from teachers about the subject content to the DfE, as the government is responsible for this. We are looking at how well GCSE assessments function with the exam boards (although our research has indicated that in relation to a number of key considerations, the qualifications function technically better than the legacy GCSEs). Changes to either might in the long run have an impact on entries in this subject.

Implementation

We are determining with exam boards the most appropriate technical approach to implementing this adjustment. In doing so we will consider a range of options and determine which will be most appropriate to adjust grading standards in GCSE French and German. However, any adjustment we propose will reflect the following four principles:

- That the approach we adopt to adjusting grading standards in French and German will not result in these GCSEs becoming statistically more severe at any grade
- We may decide to phase the implementation of the adjustment over two or more years, if a one-off adjustment would lead to undue unfairness between students in adjacent years of entry
- Having completed our adjustment to grading standards in these subjects, we do not have plans to revisit the scale of the change made in future years if the apparent 'difficulty' of these subjects under statistical measures were to shift and/or entries to the subjects did not increase
- We will seek to make changes at grades 4 and above only, but changes made at grade 4 could have a small impact on grades 2 and 3

We will provide more information on the exact adjustment that will be made, and the implications for teachers and students, ahead of awarding these qualifications in 2020.

We do not plan to revisit the decision we took in regard to grading standards in A level science or modern foreign languages. The evidence we gathered at A level was different from that at GCSE, and we have been consistent in our approach in systematically considering this evidence against our criteria.

This concludes our planned work on inter-subject comparability. Our 2016 policy decision was that we would not seek to align grading standards in all GCSE and A level subjects on a statistical basis. Having investigated grading standards in A level science and modern foreign languages, and now GCSE modern foreign languages, we are satisfied that the potential risks from our policy decision have now been addressed. However, we will keep the evidence in this area under review.

Impact Assessment

Equality Analysis

Our General Conditions of Recognition require exam boards to design assessments in such a way as to minimise any disadvantage that may be experienced by students because of their particular protected characteristics. We also require that exam boards put in place arrangements which allow reasonable adjustments to be made for disabled students when taking assessments. In this way, by the time grade boundaries are set for a particular assessment, steps have been taken to ensure that, as far as possible, all students have been given the opportunity in that assessment to demonstrate their knowledge, skills and understanding on a level playing field.

Grade setting therefore focuses on the level of the knowledge, skills and understanding which has been demonstrated in those assessments, and does not take account of the particular protected characteristics of the individual students who have taken those assessments. The grade awarded to each individual student solely reflects the performance of that student in that assessment. To do otherwise would risk introducing different standards in the same qualification for students with particular protected characteristics and those without, undermining the value of the qualification.

Given this, in respect of the adjustments we will require exam boards to make to grading standards in French and German which will apply to all those who take these qualifications, we have not identified any potential impacts on students because they share a particular protected characteristics.⁸

⁸ The protected characteristics under the Equality Act 2010 are: age, disability, gender reassignment, marriage and civil partnerships, pregnancy and maternity, race, religion or belief, and sex or sexual

Regulatory Impact and Growth

While we think that an adjustment to grading standards is unlikely to lead on its own to very significant increase in entries for GCSE French and German, we recognise some stakeholders may hold a different view. If entries for French and German do rise as a result, we believe this would be appropriate given the evidence we have considered that standards should be adjusted. We are also satisfied that exam boards would have the capacity to cope with an increase in uptake. Although there may be some small costs to exam boards to communicate with schools and colleges about the changes to be made to grading standards, we do not think that this will result in any other additional costs or burdens to them.

It is possible that part of the growth in GCSE Spanish entries over the past two decades has been the result of some schools who perceive French and German to be severely graded switching to an alternative modern foreign language. If this is the case then aligning these subjects more closely may prompt schools to change their provision, meaning any increase in entries in French and German entries comes at the expense of entries in Spanish. We believe this is unlikely however, as any schools who may have chosen to switch subjects in this manner will likely have invested significant resources in doing so (including staffing), and the evidence we have considered suggests that entries to subjects are influenced by a wide range of factors. Nonetheless, as above, if this did occur we believe this would be appropriate given the evidence we have considered that standards should be adjusted.

orientation. Under section 149 of the Equality Act 2010, we are not required to consider the impact of the reforms on those who are married or in a civil partnership.

References

- Association of School and College Leaders (ASCL). (2017). Information Paper on A Level and GCSE Modern Foreign Language Results. ASCL: Leicester. Available online at: <https://www.ascl.org.uk/download.F174118B-6F1C-43AE8DD482A055D8AD1A.html>
- Barrance R. & Elwood J. (2018) Young people's views on choice and fairness through their experiences of curriculum as examination specifications at GCSE, *Oxford Review of Education*, 44:1, 19-36, DOI: 10.1080/03054985.2018.1409964. <https://doi.org/10.1080/03054985.2018.1409964>
- Barton, G., Koglbauer, R., Davidson, S. and Richardson, W. (2017). Letter to Ofqual. Available on line at: <http://www.all-london.org.uk/site/wpcontent/uploads/2017/08/MFL-letter-to-R-Taylor-Ofqual-6-July-2017.pdf>
- Bauckham, I. (2016). Modern Foreign Languages Pedagogy Review. Available on line at: <https://www.tscouncil.org.uk/wp-content/uploads/2016/12/MFLPedagogy-Review-Report-2.pdf>
- Bramley, T. (2011). Developing Methods to Address Comparability Issues. Paper presented at the Perspectives from Cambridge Assessment Seminar 'Comparability of examination standards', 6 April 2011. London.
- Bramley, T. (2016). The effect of subject choice on the apparent relative difficulty of different subjects. Cambridge Assessment Research Report. Cambridge Assessment: Cambridge.
- Carruthers, J., Forsdick, C., Kenny, N., Kohl, K., Mentchen, S., Gratzke, M., Davies, C., Burgwinkle, B., Gorrara, C., Littler, M., Torres, I. and Honess, C. (2017). Letter to Sally Collier from Higher Education modern language subject associations.
- Churchward, D. (2019) Recent trends in modern foreign language exam entries in anglophone countries. Ofqual: Coventry.
- Coe, R. (2008). Comparability of GCSE examinations in different subjects: An application of the Rasch model. *Oxford Review of Education* 34, 609 – 636.
- Coe, R. (2010) Understanding comparability of examination standards. *Research Papers in Education* 25, 271–284.
- Coe, R., Searle, J., Barmby, P., Jones, K., and Higgins, S. (2008). Relative difficulty of examinations in different subjects. Report for SCORE (Science Community Supporting Education). CEM Centre, Durham University: Durham.
- Coffey, S. (2016). Choosing to Study Modern Foreign Languages: Discourses of Value as Forms of Cultural Capital. Oxford University Press 2016. *Applied Linguistics* 2018: 39/4: 462–480.
- Cuff, B. (2017). Perceptions of subject difficulty and subject choices: Are the two linked, and if so, how? Ofqual: Coventry. Available on line at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/614444/Perceptions_of_subject_difficulty_and_subject_choices.pdf

[achment_data/file/649891/Perceptions_of_subject_difficulty_and_subject_choices.pdf](#)

- Curcin, M., Black, B. (2019). Investigating standards in GCSE French, German and Spanish through the lens of the CEFR. Ofqual: Coventry.
- Curcin, M., Black, B. (2019). Awarder and stakeholder surveys on GCSE MFL performance standards. Ofqual: Coventry.
- Curcin, M., Black, B. and He, Q. (2018). Inter-subject Comparability – higher education representatives' perception of grade standard adjustment in some MFL and science A levels. Ofqual: Coventry.
- Dearing, R. (2007). Languages Review. Department for Education and Skills: London.
- Gill, T. (2017). The impact of the introduction of Progress 8 on the uptake and provision of qualifications in English schools. Cambridge Assessment Research Report. Cambridge, UK: Cambridge Assessment.
- He, Q., and Stockford, I. (2015). Inter-Subject Comparability of Exam Standards in GCSE and A Level. Ofqual: Coventry.
- He, Q., and Black, B. (2019). Statistical evidence pertaining to the claim of grading severity in GCSE French, German and Spanish and the impact of statistical alignment of standards on outcomes. Ofqual: Coventry.
- Jones, B. E. (2004). Inter-subject standards: An investigation into the level of agreement between qualitative and quantitative evidence in four apparently discrepant subjects. Paper presented at the annual conference of the International Association for Educational Assessment, June 13-18. Philadelphia, USA.
- Jones, B.E., Philips, D. and van Krieken, R. (2011). Inter-subject standards: An insoluble problem? AQA: Manchester.
- Kohl, K. and Mentchen, S. (2017). Letter to Richard Garrett on A level examinations and grading in modern foreign languages.
- Lindley, L., Clemens, S., Knibbs, S., Stevens, J., and Bagge, L. Ipsos MORI. (2019). Omnibus survey of pupils and their parents or carers: wave 5. Research report. Department for Education. March 2019.
- Lockyer, C., and Newton, P. E. (2015). Inter-subject comparability: a review of the technical literature. Ofqual: Coventry.
- Long, R. and Danechi, S. (2018). House of Commons Library. Language teaching in schools (England). Briefing Paper Number 07388, 19 December 2018.
- Macaro, E. (2008). The decline in language learning in England: getting the facts right and getting real, *Language Learning Journal*, 36:1, 101-108, DOI: 10.1080/09571730801988595. Available on line at: <https://doi.org/10.1080/09571730801988595>
- Mair, N (2017). Letter to Ofqual on modern foreign languages.
- Morse, A. (2016). Training new teachers. National Audit Office: London.
- Morse, A. (2017). Retaining and developing the teacher workforce. National Audit Office: London.

- Myers, H. and Blow, D. (2017). Grading in modern language A level and GCSE exams. ASCL presentation delivered at meeting with Ofqual staff.
- Newton, P. E. (2012). Making sense of decades of debate on inter – subject comparability in England. *Assessment in Education* 19, 251-273.
- Inter-subject comparability 87 Newton, P.E., He, Q., and Black, B. (2017). Progression from GCSE to A level: Comparative Progression Analysis as a new approach to investigating intersubject comparability. Ofqual: Coventry.
- Office of Qualifications and Examinations Regulation (Ofqual) (2015). GCE qualification level conditions and requirements. Ofqual: Coventry.
- Office of Qualifications and Examinations Regulation (Ofqual) (2015). Inter-subject comparability (2015 onwards). Ofqual: Coventry. Available on line at: <https://www.gov.uk/government/publications/inter-subject-comparability-2015-to-2016>
- Office of Qualifications and Examinations Regulation (Ofqual) (2015c). Inter-Subject Comparability of Exam Standards in GCSE and A Level: ISC Working Paper 3. Coventry, the Office of Qualifications and Examinations Regulation. This report was written by Qingping He (Research Fellow) and Ian Stockford (Executive Director for General Qualifications). Available on line at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/606044/3-inter-subject-comparability-of-exam-standards-in-gcse-and-a-level.pdf
- Office of Qualifications and Examinations Regulation (Ofqual) (2018). Inter-subject comparability in A level sciences and modern foreign languages. Ofqual: Coventry. Available on line at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/757841/ISC_Document_20.11.18.pdf
- Opposs, D. (2016). A policy position for Ofqual on inter-subject comparability. Ofqual Board paper 58/16. Ofqual: Coventry. Available on line at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/610111/Board_paper_-_Inter-subject_Comparability.pdf
- Panayiotou, S., Boulden, K., Newton, S. and Andersson, D. (2017). Omnibus Survey of Pupils and their Parents/Carers (Research report wave 2). Department for Education: London.
- Parrish, A. and Lanvers, U. (2018): Student motivation, school policy choices and modern language study in England, *The Language Learning Journal*, DOI: 10.1080/09571736.2018.1508305. Available on line at: <https://doi.org/10.1080/09571736.2018.1508305>
- Qualifications and Curriculum Authority (QCA). (2008). Grade standards in GCSE modern foreign languages. Qualifications and Curriculum Authority: London. Available on line at: http://www.all-london.org.uk/archive/Resources/severe_grading/qca-08-3570-Grade_standards_in_GCSE_MFL.pdf
- Qualifications and Curriculum Authority (QCA). (2008). Inter-subject Comparability Studies. Qualifications and Curriculum Authority: London.

- Stratton, T. and Zanini, N. (2019). Evaluating the impact of the introduction of reformed GCSE MFL assessments in 2018. Ofqual: Coventry.
- Taylor, R. (2015). Using the theory of planned behaviour to understand students' subject choices in post-compulsory education. Assessment and Qualification Alliance: Manchester.
- The Office for Standards in Education, Children's Services and Skills (Ofsted). (2015). Key Stage 3: the wasted years? Available online at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/459830/Key_Stage_3_the_wasted_years.pdf
- Thomson, D. (2017). Another attempt at a qualification-neutral progress 8 measure. FFT Education Datalab. School accountability. 24 January 2017.
- Tinsley, T. and Board, K. (2017). Language trends 2016/17 language teaching in primary and secondary schools in England survey report. Available on line at: https://www.britishcouncil.org/sites/default/files/language_trends_survey_2017_0.pdf
- Tinsley, T. and Dolezal, N. (2018). Language trends 2018. Language teaching in primary and secondary schools in England survey report. Available on line at: https://www.britishcouncil.org/sites/default/files/language_trends_2018_report.pdf
- Tinsley, T. (2019). Language trends 2019. Language teaching in primary and secondary schools in England survey report. Available on line at: <https://www.britishcouncil.org/sites/default/files/language-trends-2019.pdf>
- Tracy, C. (2016). Grading severity, choice and uptake. Presentation delivered at the Ofqual inter-subject comparability conference in February 2016 held in London.
- William, D. (1996). Meaning and consequences in standard setting. *Assessment in Education* 3, 287–307.
- Wyburd, J. (2016). Modern Languages from school to university and beyond: the current landscape. Available online at: <https://www.ismla.co.uk/uploads/1492437369-ISMLA2016JW.pptx>



© Crown Copyright 2019

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

To view this licence, visit

www.nationalarchives.gov.uk/doc/open-government-licence/

or write to

Information Policy Team, The National Archives, Kew, London TW9 4DU

Published by:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

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