

ANNEX A

The case for adjusting A level grading standards in MFL

Response to Policy Decision *Inter-subject comparability in A level sciences and modern foreign languages. Examining the claim that these subjects are more severely graded than other A levels* (Ofqual, Nov. 2018)¹

The inter-subject comparability study was undertaken by Ofqual on the basis of the decision to ‘adjust grading standards in individual subjects where there was an exceptional and compelling case’ and to ‘begin by looking at A levels in physics, chemistry and biology, and French, German and Spanish’ (p. 3).² Ofqual concludes ‘that there is not a compelling case to adjust grading standards in A levels’ in these subjects (p. 5).

The paper presented here in response to the Policy Decision sets out the reasons for instead considering MFL to be an exceptional case and for adjusting grading standards in French, German and Spanish, and in the other modern foreign languages in which A levels are offered.

EXECUTIVE SUMMARY

Ofqual’s criteria

Four criteria are used by Ofqual in the study, on the grounds ‘that no single piece of evidence could definitively demonstrate the case for an adjustment to grading standards in a given subject’ (p. 6):

- a) Statistical measures of subject difficulty show evidence of persistent grading severity over several years
- b) There is persuasive evidence of the potential detrimental impact caused by severe grading on those who use the qualification and on society at large over several years
- c) There is evidence which shows that those who use the qualification and those responsible for maintaining the grading standard judge an adjustment to be acceptable
- 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 (pp. 8, 9, 12, 15).

It is not made clear what strength of fulfilment is required severally or jointly. This is a concern especially in light of the following:

- The common denominators for the subjects included in the study are that they are ‘of above average difficulty’ (pp. 8f.) and that they are subject to ‘claims they are more severely graded than other A level subjects’ (p. 3). This statistical evidence (criterion (a)) unequivocally supports the claim of severe grading, whereas the evidence provided for criteria (b)–(d) is less well founded.
- The conclusions (‘Our view’, pp. 20f.) appear to assume that all criteria must be fully met in all subjects considered to prompt an adjustment of grading. This has no basis in the earlier part of the report.
- No justification is provided for coupling MFL and sciences with respect to criteria (b) – (d).
- Little evidence is provided in the Policy Decision document for the impact on society (criteria (b) and (d)) with respect to MFL, with evidence in the public domain compiled by organisations other than Ofqual playing no part in the explanations. A detailed discussion of such evidence is provided in the Technical Report, with the concluding focus that there is insufficient evidence of causal links between severe grading and factors impacting negatively on the health of the subject.³

¹ <https://www.gov.uk/government/publications/inter-subject-comparability>. Page references refer to this document unless otherwise indicated.

² Physics, chemistry, biology, French, German and Spanish each receive separate comment in the report and are for the most part treated individually but the title of the study refers to ‘sciences and modern foreign languages’ and these collective concepts are also used occasionally elsewhere, notably in the section ‘Our view’, p. 21.

³ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/757840/Inter_subject_comparability_-_technical_report_MFL.pdf, pp. 23-50.

Investigating grading in MFL together with grading in sciences

The rationale for investigating sciences and MFL in a single study is that the research was undertaken ‘in response to claims they are more severely graded than other A level subjects’ (p. 3). The statistical evidence presented in the report yields the following degrees of overall severity for the years 2013 and 2017, with all 6 subjects being ‘of above average difficulty’ (pp. 8f.):

Physics (2nd); 3rd: Chemistry (3rd); Biology (4th/5th); French (7th); German 8th); Spanish (9th/13th).

Notwithstanding the term ‘inter-subject comparability’, the six subjects are mostly treated as distinct in the main body of the report, and there is no statement to the effect that conclusions reached for one set of subjects must also apply to the other. The report refers to Ofqual’s earlier decision ‘not to take co-ordinated action to align grading standards across all GCSE and A level subjects according to statistical measures of subject difficulty’, and the decision potentially to adjust grading standards ‘in *individual* subjects where there [is] an exceptional and compelling case’ (p. 3, underlining added).

The conclusions set out in the final section ‘Our view’ (pp. 20f.), however, suggest that the conclusions reached for the sciences influenced or determined the conclusions reached for MFL:

- i. The section ‘Our view’ draws analogies between MFL and sciences which suggest – without justifying the assumption – that the two groups of subjects can be automatically aligned with regard to causes and effects of grading anomalies (p. 20).
- ii. The same section asserts a ‘logical’ connection between potential action taken for MFL and action taken for sciences (p. 21), evidently based on the assumption that the parameters affecting grading in MFL and sciences are identical.
- iii. For criterion (b), the report concludes with respect to MFL that evidence ‘is apparently strong, but causation is questionable’ (p. 19). The doubt cast on a causal connection between severe grading and a negative impact on exam entries is supported by reference to Spanish, where gradually rising entries from 2008 to 2017 are ‘bucking the trend’ (p. 11): ‘A number of universities have stopped offering single and joint honours degrees in the language over the past decade. The fact that this has happened in Spanish despite increasing A level entries may call into question the assertion of stakeholders that the negative trends in this subject are attributable in any significant way to the effects of severe grading’ (p. 12). Similarly: ‘Apparently persuasive evidence under criterion b in the form of declining A level entries for [French and German] becomes less convincing when considered alongside Spanish’ (p. 20; see also 20f.).
The use of evidence concerning university provision specifically in Spanish to negate a causal connection between severe grading and plummeting exam entries in MFL overall is fundamentally flawed in assuming that developments in one language in a university department are unaffected by developments in other languages; that a university can or will respond automatically to changes of trend feeding through gradually from schools; and that the provision of university courses can be viewed in isolation from factors such as the respective university’s academic and financial policy.
Ofqual’s in-house speculation about higher education provision is here being used as evidence to justify leaving severe grading unaddressed.

Differences between MFL and sciences relevant to the criteria

The comments cited under (i.)–(iii.) suggest that Ofqual has taken no account of key differences between MFL and sciences, most importantly the relevance of optionality at GCSE: MFL was made optional in 2004 while all three sciences are compulsory at GCSE. By comparison with sciences, where severe grading at GCSE has no impact on GCSE numbers, and a large pool is available for A level, MFLs reputation as a ‘difficult’ subject discourages take-up already at GCSE, causing a direct impact on provision in schools and higher education, on teacher training, and on society as a whole. The difference is evident in the number of entries:

Number of entries in 2018: Fre/Ger/Spa: – c. **256k** at GCSE, c. **19k** at A level
(Compare:) Bio/Chem/Phys/Comb: c. **1,230k** at GCSE, c. **145k** at A level

Further critical differences include the differing academic profile of the MFL and sciences cohorts given the impact of self-selection at GCSE, and especially the participation of native and near-native speakers in MFL exam cohorts. An additional critical factor is the massive investment in UK sciences over the past decade and the high status of the subject in UK society, which can counteract effects of severe grading and allow this to be associated with prestige especially for high achievers. By contrast, MFL has been beset by policy failures, under-investment and low prestige in a society that increasingly considers language skills unnecessary.

Evidence for assessing whether the case to adjust A level grading standards in MFL is compelling

- The statistical evidence reported under criterion (a) shows that French, German and Spanish are ‘of above average difficulty’. No justification is provided for assuming that above average difficulty in statistical terms (criterion a) should be addressed with a grading adjustment *only* if criteria (b), (c) and (d) are *all* fulfilled, or addressed only if there is also a justification for a grading adjustment in sciences.
- In the Policy Decision document, Ofqual’s conclusions concerning criteria (b) and (d) are supported neither with existing research in the public domain nor with appropriate expertise sought in the context of the study. Evidence is discussed in the Technical Report, with a focus on highlighting absent causal links between severe grading and falling numbers of entries. This does not however take account of the impact of severe grading on learner experience, confidence and motivation, or of the fact that a subject which impacts negatively on a learner’s grade profile for university applications will not be convincing as an attractive choice for the learner, their parents or their school, especially if other factors are also impacting negatively on the subject.
- The report in effect ignores the key issue of the participation of native and near-native speakers in MFL exam cohorts. It refers to this only in references to the one-off adjustment made by Ofqual in 2017 and in the occasional comment suggesting a lack of appreciation that this remains an ongoing key issue concerning grading in MFL and one that is not found in any other subject. Notwithstanding the fact that Ofqual acknowledged the challenges of quantifying it in 2017, and recognised its statistical significance, no provision is envisaged in the report for monitoring or further addressing this factor. This is particularly critical for German, where the adjustment was wholly inadequate in addressing the finding that at A*, almost half the students in the sample were native-speakers, and at grade A, almost a fourth.⁴ This finding was evidently not taken into account in the Policy Decision, given the comment that ‘German generally [appears] to be more lenient than the sciences and French’ (p. 9).

It is notable that the Technical Report (pp. 72-77) indicates many reservations about a grade adjustment on the part of the exam boards. Their views appear in fact to have played a more significant role in Ofqual’s decision-making and conclusions than is obvious from the Policy Decision document or than one might have expected given their natural interest in keeping the status quo, lack of expertise with respect to the social impact of severe grading, and lack of first-hand knowledge of university selection procedures.

On the basis of the evidence provided in the Policy Decision report, it is reasonable to conclude that all four criteria defined by Ofqual as relevant are met for MFL.

Overall, the evidence set out in the Policy Decision document indicates that the case to adjust A level grading standards in MFL is compelling.

⁴https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/610050/Native_speakers_in_A_level_modern_foreign_languages.pdf, p. 37.

RESPONSE IN FULL

Ofqual's criteria

Four criteria are used by Ofqual in the study, on the grounds 'that no single piece of evidence could definitively demonstrate the case for an adjustment to grading standards in a given subject' (p. 6).

[See the Executive Summary for the four criteria given in the report.]

It is potentially persuasive that criteria beyond statistical measures should be taken into account when addressing the issue of appropriateness of grading standards, and that criteria relating to the use of qualifications by stakeholders and their role in society should be included in the 'basket of evidence'. However, the appropriateness and weight of these criteria must depend on the way they are used.

It is a concern that the status and weight of the four criteria in relation to each other is not addressed, and it is not made clear what strength of fulfilment is required severally or jointly. This is a concern especially in light of the following:

- The common denominators for the subjects addressed in the study are that they are 'of above average difficulty' (pp. 8f.) and that they are subject to 'claims they are more severely graded than other A level subjects' (p. 3). This statistical evidence (criterion (a)) unequivocally supports the claim of severe grading, whereas the evidence provided for criteria (b)–(d) is less well founded.
- In the final part of the report, which sets out Ofqual's conclusions ('Our view', pp. 20f.), the 'basket of evidence' appears to become an inventory to be completed across all the subjects included in the study if the evidence is to prompt an adjustment of grading. This has no basis in the earlier part of the report.
- No justification is provided for using grading specifically in sciences as a criterion for decision-making with respect to grading in MFL beyond the fact that they are statistically 'of above average difficulty' (pp. 8f.) and subject to claims of severe grading.
- Little attempt is made to produce evidence beyond speculation for the impact on society (criteria (b) and (d)) with respect to MFL. There is no indication that evidence compiled by other organisations available in the public domain has been taken into account.

Concerning criterion (c), it is persuasive that Ofqual should have consulted examiners and HE representatives. One might however have expected schoolteachers also to be formally consulted. Peculiarly, the following selective reference is made to schoolteachers with respect to Physics: 'The view that teachers were generally satisfied with standards in this subject was expressed' (p. 13). One might then expect mention of the evidence provided by an ASCL survey of MFL teachers in November 2017, which generated over 2,600 responses in under a week urging Ofqual to address severe grading.⁵ Moreover, parts of the report suggest that the views of representatives from the exam boards were given excessive weight in some areas, notably with respect to selection criteria for HE courses and impact on society (see e.g. pp. 15-17 concerning criterion (d)).

Investigating grading in MFL together with grading in sciences

The rationale for investigating sciences and MFL in a single study is that the research was undertaken 'in response to claims they are more severely graded than other A level subjects' (p. 3). The statistical evidence presented in the report yields the following degrees of overall severity for the years 2013 and 2017, with all 6 subjects being 'of above average difficulty' (pp. 8f.):

Physics (2nd); 3rd: Chemistry (3rd); Biology (4th/5th); French (7th); German 8th); Spanish (9th/13th).

Notwithstanding the term 'inter-subject comparability', the six subjects are mostly treated as distinct in the main body of the report, and there is no statement to the effect that conclusions reached for one set of subjects must also apply to the other. This would indeed be out of line both with the earlier decision 'not to take co-ordinated action to align grading standards across all GCSE and A level subjects according to statistical

⁵ 'Survey on GCSE ML severe grading', published 29 November 2017, <http://www.all-london.org.uk/site/wp-content/uploads/2017/11/Survey-on-GCSE-ML-severe-grading-Nov-17-final-doc.pdf>.

measures of subject difficulty', and with the decision potentially to adjust grading standards 'in *individual* subjects where there [is] an exceptional and compelling case' (p. 3, underlining added). In discussing potential conclusions (Options A-D, pp. 19f.), the two groups of subjects are treated separately.

The conclusions set out in the final section 'Our view' (pp. 20f.), however, suggest that the conclusions reached for the sciences influenced or determined the conclusions reached for MFL:

- i. The section 'Our view' draws analogies between MFL and sciences which suggest – without justifying the assumption – that the two groups of subjects can be automatically aligned with regard to causes and effects of grading anomalies:
'French and German both appear to be more lenient under the various statistical measures of subject difficulty than physics, chemistry and biology – all of which are experiencing an increase in entry overall' (p. 20).
- ii. The same section goes on to assert a 'logical' connection between potential action taken for MFL and action taken for sciences: 'If we had decided that severe grading was having an impact upon uptake in modern foreign languages, the extent to which those subjects appear more lenient than physics, chemistry and biology means that logically we should also make an adjustment to grading standards in the sciences' (p. 21). What is presented here as a 'logical' connection in effect assumes that the parameters affecting grading in MFL and sciences are the same.
- iii. For criterion (b), the report concludes with respect to MFL that evidence 'is apparently strong, but causation is questionable' (p. 19). The doubt cast on a causal connection between severe grading and a negative impact on exam entries is supported by reference to Spanish, where gradually rising entries from 2008 to 2017 are 'bucking the trend' (p. 11):
'A number of universities have stopped offering single and joint honours degrees in the language over the past decade. The fact that this has happened in Spanish despite increasing A level entries may call into question the assertion of stakeholders that the negative trends in this subject are attributable in any significant way to the effects of severe grading' (p. 12).
Similarly: 'Apparently persuasive evidence under criterion b in the form of declining A level entries for [French and German] becomes less convincing when considered alongside Spanish' (p. 20; see also 20f.).
The use of evidence concerning university provision specifically in Spanish to negate a causal connection between severe grading and plummeting exam entries in MFL overall is fundamentally flawed in assuming that developments in one language in a university department are unaffected by developments in other languages; that a university can or will respond automatically to changes of trend feeding through gradually from schools; and that the provision of university courses can be viewed in isolation from factors such as the respective university's academic and financial policy. Ofqual's in-house speculation about higher education provision is here being used as evidence to justify leaving severe grading unaddressed.

In this context it is also notable that the press release announcing publication of Ofqual's Policy Decision on severe grading introduces the concept of a 'uniformly compelling case' without making clear whether this means 'uniform' across criteria or across subjects, and if the latter, whether it relates to uniformity within each of the six subjects, within sciences and MFL respectively, or across the six subjects investigated:

Adjusting grading standards is something we would only consider if there were a uniformly compelling case to do so. We have carefully considered the arguments of stakeholders and closely examined a broad range of evidence against our criteria. We have concluded there is no such uniformly compelling case to adjust grading standards in these subjects.⁶

⁶ Ofqual press release 'Reviewing inter-subject comparability', 21 November 2018, <https://www.gov.uk/government/news/reviewing-inter-subject-comparability>, p. 2.

The word 'uniform' is not used in the Policy Decision report, and there is nothing in the definition of the principles and criteria relevant to the study to suggest that any action resulting from conclusions reached for the sciences has bearing on action resulting from conclusions reached for MFL. On the basis of the principles and criteria, it ought to be assumed that any conclusions reached for MFL French, German and Spanish are valid and sufficient as a basis for determining any actions with respect to adjusting grading standards, irrespective of findings for sciences.

Differences between MFL and sciences relevant to the criteria

It is vital to recognise the significant differences between sciences and MFL with respect to the criteria used by Ofqual in its study. These differences are ignored at certain points in the study and not taken into account in the conclusions. Indeed the comments cited under (i.)–(iii.) suggest that Ofqual has taken no account of key differences between MFL and sciences, most importantly the relevance of optionality at GCSE.

- **Optionality at GCSE:** MFL was made optional in 2004 while all three sciences are compulsory at GCSE. By comparison with sciences, where severe grading at GCSE has no impact on GCSE numbers, and a large pool is available for A level, MFLs reputation as a 'difficult' subject discourages take-up already at GCSE, causing a direct impact on provision in schools and higher education, on teacher training, and on society as a whole.

The impact of this factor with respect to MFL is consequently significant for criteria (b) and (d), and differentiates MFL significantly from sciences with respect to those criteria.

Optionality allows schools to reduce or cut provision of MFL already at GCSE, an effect that especially affects schools which are under financial pressure, and under-performing schools. This means that above-average severity of grading can cause schools to drop MFL altogether at GCSE, and it has over many years contributed significantly to loss of critical mass from GCSE through to tertiary level. This factor is distinctive for MFL and it has high and distinctive relevance for criteria (b) and (d).

- **Academic profile of the cohorts:** MFL is commonly seen as a subject for academically able students, which means that lower performing students are likely to opt out of their own accord or be discouraged by schools and parents from choosing MFL. This feature of MFL cohorts ought not to have an effect on statistical measures concerning fairness of grading using Rasch analysis and Comparative Progression Analysis. However, it is likely to have driven up the difficulty of the exam papers over time, in conjunction with the participation of native and near-native speakers, increasing the (justified) perception of MFL as a 'difficult' and consequently unattractive subject.

- **Number of entries:** There is a very significant difference between entry numbers in MFL and sciences:

- GCSE French, German, Spanish in 2018 overall: **255,845**
- GCSE Biology, Chemistry, Physics, Combined science in 2018 overall: **1,229,665**
- A level French, German, Spanish in 2018 overall: **18,580**
- A level Biology, Chemistry, Physics in 2018 overall: **145,010**

This difference is caused by a wide range of factors including optionality, investment and status within UK society, with cumulative effects that must be taken into account in any comparison between these subjects with respect to the criteria. It means that any negative impact of above-average severity of grading is immediately more critical for MFL than for sciences.

- **Trend:** While entries in science A levels are buoyant, entries in MFL are in decline overall (pp. 10f.):
 - Physics 8th most popular A level in 2018, with rising entries; Chemistry high entries; Biology 2nd most popular subject.
 - French: long-term decline, from 15,000 to under 8,000 in past decade; German: decline as for French, down to under 3,000; Spanish gradual increase since 2008, slight downturn in 2018.This factor is relevant to (b) and (d).

- **Native and near-native speakers:** The participation of students with native-speaker competence in every MFL cohort was recognised by Ofqual as relevant to grading standards in a small one-off adjustment of grading.⁷ This factor distinguishes MFL from all other academic subjects as does the participation of students with near-native-speaker advantage. The complexity of identifying the impact of native speakers on grading are considerable, and Ofqual acknowledges that it is not possible to take full and reliable account of this factor statistically, or at all for languages other than French, German and Spanish. It is also clear that the adjustment did not go nearly far enough for German, where an Ofqual study published in 2017 found that almost half the A* students in the sample were native speakers while almost a fourth of grade A students were native speakers.⁸ Moreover, as entries from candidates with English as their first language decreases, it is likely that the proportion of native speakers increases, given that they are the least likely group to drop out of the subject.
- **Investment and status within UK society:** The past decade has seen government, academies, trusts and subject associations seeking to boost take-up and progression in schools with considerable investment. The Green Paper *Building our Industrial Strategy* (2017) sets out a strategy for strengthening STEM subjects while languages are not mentioned. Accordingly, sciences are widely perceived to offer significant career and status advantages while languages are not valued in this way. These differences are relevant to the criteria. For example, they help to explain why progression to A level is proportionately higher in sciences than in MFL notwithstanding the fact that there is no academic selection or self-selection for sciences at GCSE, and notwithstanding any perceptions of severe grading. In conjunction with the far higher numbers of students and greater health of the sciences across educational sectors, they also explain why there is less concern in the sciences about a potential detrimental impact of severe grading on users of the qualification and society. .

The differences between sciences and MFL outlined above are all significant with respect to the criteria used to identify whether there is a compelling case for an adjustment to grading standards, and they interact in ways that are significant.

Evidence for assessing whether the case to adjust A level grading standards in MFL is compelling

The key question at the centre of evaluating the evidence presented by Ofqual to justify its Policy Decision is whether there is ‘a compelling case to adjust A level grading standards in MFL’ or whether ‘Option A’ is justified: ‘Take no action [...] on the basis that our criteria for a compelling case have not been met’ (p. 19). This question will be considered in the following on the basis of the evidence provided for the four criteria.

CRITERION (a) (pp. 8f.)

- According to Ofqual’s findings, French, German and Spanish are all ‘of above average difficulty’ (p. 8).
- The report refers to the fact that Ofqual has recognised the significance of the participation of native and near-native speakers on grading and made a one-off adjustment in 2017 (p. 9). However, it does not mention the fact that Ofqual acknowledges the difficulty of establishing the extent of the impact (see above, note 4). It also does not mention that the participation of native speakers in German is recognised to be exceptionally high, especially at the top end of the scale (‘at A* [...] almost half of the students in our sample are native-speakers, and at grade A [...] almost a fourth of the students are native speakers’,

⁷ See p. 9. For a discussion of this issue see the publication cited in footnote 3. In 2016, the percentage of native speakers in A-level cohorts based on responses from teachers was found to be as follows. All students in the 3 A level MFLs: French 8.7%, German 17.4%, Spanish 10.1%. Percentages for ‘matched’ students in the 3 A level MFLs (i.e. those matched to their prior GCSE attainment and therefore included in grading statistics): French 7.6%, German 10.5%, Spanish 4.9% (ibid., p. 34). The article makes clear, and Ofqual has repeatedly acknowledged, the difficulty of establishing a robust basis for appropriate adjustment, and the likelihood of variation over time.

⁸ See the publication cited in footnote 3, p. 37. The research makes clear that there are significant differences between languages with respect to the effect: ‘For French and Spanish, the proportion of native speakers achieving each grade (of the overall number of students) is relatively constant, yet for German, the situation is different: in percentage terms, there are more native speakers achieving the top grades than those achieving lower grades (ibid.).’

see above, note 5). The statement in the Policy Decision report that ‘German generally [appears] to be more lenient than the sciences and French’ (p. 9) fails to take account of the difference between native speakers and non-native speakers – for the latter, the highest marks are likely to be significantly more difficult to achieve than average. makes clear that important available research used by Ofqual in other contexts was ignored when assessing the statistical measures of difficulty in MFL.

- ‘Option A’: The basis for the conclusion that ‘there is a lack of persuasive evidence for criterion a’ (p. 19) is not transparent.

On the basis of the evidence provided in the report, this criterion is met for MFL.

CRITERION (b) (pp. 9-12) In order to judge whether there is ‘persuasive evidence of the potential detrimental impact caused by severe grading on those who use the qualification and on society at large over several years’, Ofqual expects to see evidence such as the following:

- i. Depressed uptake within the courses to which students taking the subject would be expected to progress
- ii. Depressed entries within the subject
- 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 qualification.’ (pp. 9f.)

- The figures and factors listed for French and German (p. 11) give a picture of catastrophic decline, while offering no comment on any potential connection with severe grading.
- Concerning Spanish, where entries are ‘bucking the trend’ (p. 11), the report makes the following point: ‘A number of universities have stopped offering single and joint honours degrees in the language over the past decade. The fact that this has happened in Spanish despite increasing A level entries may call into question the assertion of stakeholders that the negative trends in this subject are attributable in any significant way to the effects of severe grading.’ (p. 12)

This assertion of causal connections is fundamentally flawed above all in assuming that developments in one language in a university department are unaffected by developments in other languages, or that such developments can be viewed in isolation from factors affecting the specific department in the context of strategic, academic and financial policy within the specific university.

The reference to ‘the negative trends in this subject’ is opaque since the focus concerning Spanish is on a positive trend (see criterion (a)), albeit with a slight downturn noted for 2018 (p. 11).

- There is no reference in this section to the ample evidence of negative impacts relating to (i) – (iv) that is available in the public domain. Evidence is discussed in the Technical Report, with a focus on highlighting absent causal links between severe grading and falling numbers of entries. This does not however take account of the impact of severe grading on learner experience, confidence and motivation, or of the fact that a subject which impacts negatively on a learner’s grade profile for university applications will not be convincing as an attractive choice for the learner, their parents or their school, especially if other factors are also impacting negatively on the subject.
- In considering ‘Option A’ (‘Take no action [...] on the basis that our criteria for a compelling case have not been met’, p. 19), it is stated that ‘the evidence under criterion b is apparently strong, but causation is questionable’. The report fails to make clear that no evidence is provided with respect to causation. There is no indication of in-house research or scrutiny of recent relevant reports by organisations such as the British Council and the British Academy. There is no mention of a causal connection for French and German, and for Spanish, there is only ill-founded speculation about a lacking causal connection.

On the basis of the evidence provided in the report, this criterion is met for MFL.

CRITERION (c) (pp. 12-15)

- There is a marked difference between HE responses for sciences and MFL. While in sciences, HE participants did not generally favour an adjustment, support for this was consistent in MFL. For French, ‘There was strong support from participants for an adjustment to grading standards’ esp. at A/B (p. 13); for German, ‘support for adjusting grading standards was strong, with one participant claiming the majority of scripts were awarded one grade lower than they deserved’ (p. 14); for Spanish, there was ‘very strong support from participants for lowering grade thresholds at AB, and more substantial adjustments at A*/A and C/D. Unanimous support for an adjustment of over five marks at the B/C threshold. [...] This was the strongest support for an adjustment seen in the research study’.

- Strong support in MFL is also stressed in the relevant further report:⁹ ‘the script review exercise suggested that there may be less appetite for grade standard adjustment in the sciences, in contrast with the languages.’ ‘More often than not in the science subjects, the discussion by the panel indicated a lack of acceptance of any adjustment [...].The opposite general pattern was apparent in the languages, where, even in the few cases when the outcome of the script review suggested a lack of clear acceptance of grade standard adjustment, the discussions were overwhelmingly in favour of adjustment.’
- The above does not support the claim under ‘Option A’ (‘Take no action [...] on the basis that our criteria for a compelling case have not been met’, p. 19) that ‘The evidence under criteria d and to some extent c is mixed’.

On the basis of the evidence provided in the report, this criterion is met for MFL.

CRITERION (d) (pp. 15-17)

- The summary (pp. 15f.) appears to be based only on findings relevant to the sciences. In particular, references to the risk of increased university entry requirements have no relevance for MFL.
- There is some overlap and misalignment between the findings reported under this criterion and those reported under Criterion (c).

For sciences, both HE and exam boards are shown to be aligned in not favouring an adjustment to grading standards.

For French, HE participants were in favour of an adjustment at A/B (but see also p. 13 re favouring adjustment at B/C). The report here dwells mainly on views and concerns by exam boards, which ‘disagreed over whether an adjustment to standards would be necessary to address the decline in entries in French’, and speculated that ‘lowering standards at grade A would likely make it more challenging for universities to identify students with the necessary grammatical knowledge to cope with undergraduate courses’. This was not supported by HE.

For German, the report here suggests very lukewarm support from HE and exam boards, yet strong support is outlined under Criterion (c).¹⁰ Speculation by exam boards (‘intervention to adjust grading standards might be unnecessary’) is given excessive prominence here as for French.

For Spanish, strong support from HE, less unequivocal support from exam boards.

Overall, HE is shown to be supportive of adjusting grading standards, especially if account is also taken of what is reported on the same issue under Criterion (c).

- The report on the workshop states that ‘The participants across all 3 panels were generally of the opinion that adjusting grade standards would not have a negative impact and that they would have been more than happy to accept students into their institutions who just missed the required grade.’¹¹

⁹https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/757837/Inter_subject_comparability_HE_perceptions.pdf, see pp. 4f.

¹⁰ Dr Joanna Neilly (University of Oxford), who participated in the relevant workshop for German on 7-8 February 2018, reported as follows: ‘There was very broad consensus that:

- The model borderline scripts should not have been borderline at all, but clearly within the higher grade. (this was not one of the study’s questions; nonetheless many of us raised the point as an important one). One of the model A* scripts, which was only just an A*, was excellent. To put it into context, if I had a first-year undergraduate at the end of the first year who produced the written work this candidate had done, I would be very impressed. A colleague from another institution thought that this written work would pass the final-year university exam, perhaps at a solid 2.ii level.
- The grade boundaries should be lowered. Only one panellist had some reservations about this, because she was from an institution that often accepts C grades even though their official policy is that a B is required. She was therefore concerned about getting students who might be of a D-grade standard, if the boundaries were lowered. (My own view was that, although the D scripts and orals were evidently weaker than the others, only one of them, out of about 6 or 7, was really deserving of a D and the others should have been low Cs.)’

¹¹https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/757837/Inter_subject_comparability_HE_perceptions.pdf, p. 29. See also the comments on pp. 30f.

- The above does not support the claim under ‘Option A’ (‘Take no action [...] on the basis that our criteria for a compelling case have not been met’, p. 19) for MFL from the point of view of HE that ‘The evidence under criteria d and to some extent c is mixed’.

On the basis of the evidence provided in the report and supporting documentation, this criterion is met for MFL at least with respect to HE. Concerns voiced by exam boards are based to a considerable extent on speculation about entry to higher education which was not supported by HE representatives.

OFQUAL’s Conclusions

- In its conclusions, the report inappropriately considers factors in isolation from each other for purposes of comparison between MFL and sciences, and evidence provided for the conclusions is frequently not robust.

An example is the following justification for not adjusting grading: ‘French and German both appear to be more lenient under the various statistical measures of subject difficulty than physics, chemistry and biology – all of which are experiencing an increase in entry overall. The number of universities offering joint and single honours languages has also decreased for Spanish, despite increasing A level entries. This raises questions that the negative trends in this subject are attributable to severe grading’ (pp. 20f.). This conclusion builds on a point made earlier in the report concerning Spanish, under Criterion (b): ‘A number of universities have stopped offering single and joint honours degrees in the language over the past decade. The fact that this has happened in Spanish despite increasing A level entries may call into question the assertion of stakeholders that the negative trends in this subject are attributable in any significant way to the effects of severe grading’ (p. 12). This conclusion is fundamentally flawed above all in assuming that developments in one language in a university department are unaffected by developments in other languages, or that such developments can be viewed in isolation from factors affecting the specific department in the context of policy within the specific university. It is highly problematic that unfounded speculation of this kind is presented as part of Ofqual’s decision-making on a fundamental matter of policy that claims to be based on evidence.

- A comment in the report concerning Spanish suggests that the relevance of native speakers participating in exam cohorts for the issue of grading standards was poorly understood by the awarders from the exam boards and indeed by the compilers of the report: ‘All of the panels felt that Spanish was considered to be more difficult than other subjects by students, particularly in terms of achieving grades A* and A, and that this was leading them to study alternative A levels which they considered ‘easier’. This was attributed to the impact of native speakers within the cohort, rather than a misalignment of standards’ (pp. 14f.). What seems not to be understood here is that a significant participation of native speakers within the cohort – of which this statement provides proof – is a clear indication of a real or perceived misalignment of standards *for students who do not have a native-speaker advantage* and who should be the reference point for grading. The students referred to appear to have had a better grasp of this connection than the awarders.
- It is notable that the Technical Report (pp. 72-77) indicates many reservations about a grade adjustment on the part of the exam boards. Their views appear in fact to have played a more significant role in Ofqual’s decision-making than is obvious from the Policy Decision document or than one might have expected given their natural interest in keeping the status quo, lack of expertise with respect to the social impact of severe grading, and lack of first-hand knowledge of university selection procedures.

Section ‘Our view’ (pp. 20f.)

This section is poorly argued and reaches a conclusion that is not in accordance with the findings presented under the individual criteria. As indicated above, the reference to Spanish university courses in isolation from other factors, repeated here as critically significant (p. 20f.), cannot be considered robust evidence.

- The conclusion on the way forward comes as a surprise since it is not based on findings in the preceding report:
 ‘Whilst we did not find compelling evidence to lower grading standards on the basis of inter-subject comparability, we do think limited action to address stakeholder concerns that these subjects could become more severely graded in the future is appropriate. Therefore, we have concluded that we should not make an adjustment to lower grading standards in subjects. Whilst we have however decided that we should act in relation to concerns of stakeholders that the apparent difficulty of these A levels might become more pronounced in the future.’ (p. 21).
 It should have been made clear here that the evidence for lowering grading standards in MFL is strong overall on the basis of the criteria (a) – (d) but that Ofqual evidently decided to override this in the light of the divergent findings for sciences, and the caveats by the exam boards set out in the Technical Report (pp. 72-77).
 Since the studies that formed the basis for the report focused on grading in past examinations, reference to stakeholder concerns about the future appear spurious, and would appear primarily to reflect the perspective of the exam boards. For MFL, stakeholder concerns focus on the documented existing anomalies, and using severe grades as the basis for future grading must automatically entail the continuation of severe grading.
- The conclusion to implement ‘a one-sided reporting tolerance’ is welcome, but it does not address the perceived or actual past and current anomalies and simply confirms that the anomalies will continue at the current level.

The conclusions of the report overall add up to the fact that even though grading in MFL has been statistically shown by Ofqual to have been severe in the past, and was perceived to be severe by the representatives from higher education participating in the study, Ofqual proposes nonetheless to use current standards as the baseline for grading in the future.

This is highly problematic in light of the long-standing evidence of the perceived and actual ‘difficulty’ of MFL, evidenced in a recent BBC survey of secondary schools: 76% of respondents in England cited ‘perceptions that the course/exams are too difficult’ as the factor with the biggest negative impact on MFL provision.¹² It is Ofqual’s responsibility to address this problem, in conjunction with the exam boards (see also Annex B). In this context it is important to attend to the purpose of the reformed A level in modern foreign languages: ‘to produce a rich and rewarding qualification, with an appropriate level of cognitive challenge and suitable for progression to university study or to employment’.¹³ Teachers who have to focus on instilling exam strategies do not have the scope to teach a deep understanding of the language being learned or to make the learning experience rich and rewarding. And learners who are faced with a dry diet of exam skills and find themselves failing in MFL by comparison with their other subjects will neither develop a confident and creative approach to cognitive challenge in languages nor feel encouraged to pursue their study of languages at university.

¹² BBC survey of secondary schools on language learning in the UK with response rate of over 50% (27 February 2019), <https://www.bbc.co.uk/news/education-47334374> (accessed 15.2.2019); additional data courtesy of Clara Guibourg.

¹³ The A Level Content Advisory Board, ‘Report of the ALCAB Panel for modern foreign and Classical Languages’, <https://alevelcontent.files.wordpress.com/2014/07/alcab-report-of-panel-on-modern-foreign-and-classical-languages-july-2014.pdf>, p. 2.

RECOMMENDATIONS

- 1. Revisit the evidence concerning severe grading with respect to MFL A level.**

If the evidence cited in Ofqual's Policy Decision document is interpreted appropriately, there is ample evidence demonstrating that an adjustment is appropriate. In revisiting the evidence, spurious views such as those cited for criteria (b) and (d) should be ignored, and conclusions about severe grading in MFL should be uncoupled from conclusions about severe grading in sciences. The other recommendations presented here should also be taken into account.
- 2. Uncouple severe grading concerns in MFL from severe grading concerns in sciences.**

It is essential to consider MFL with reference to the specific parameters that affect the subject and its role in society. These parameters differ from those governing the sciences.
- 3. Involve the full range of appropriate expertise and relevant evidence in investigating severe grading.**

Ofqual should systematically involve the expertise of schoolteachers, headteachers and schools-related organisations in its research and draw systematically on the wealth of evidence provided by organisations including ASCL, ISMLA and ALL. It is also vital to draw on the extensive evidence provided by the British Council, British Academy and others relevant to grading and its impact on society.
- 4. Address the impact of native speakers and near-native speakers on assessment appropriately.**

Ofqual acknowledged in 2017 that the participation of native speakers and near-native speakers has an impact on grading, and it went some way towards addressing this for French, German and Spanish with a one-off adjustment. It is essential for Ofqual to gain an understanding of this issue across languages, and differentiating between languages. Data on language qualifications taken in individual languages ahead of the normal qualification year provides one indicator for the proportion of candidates with likely native or near-native speaker advantage. Ofqual should implement an adjustment for all languages, take account of differences between languages, and monitor the impact over time. Examinees and stakeholders should not bear the burden of challenges involved in quantifying and monitoring this factor.
- 5. Investigate and address excessive linguistic challenge in MFL A level examinations.**

Fairness of grading must be investigated in conjunction with an investigation of the difficulty of MFL exams (see Annex B). It is essential for the exam papers to be appropriate to the standard of competence that can reasonably be expected from learners who have no native or near-native speaker advantage.
- 6. Address evidence of the actual and perceived 'difficulty' of MFL exams, and of the school subject as a whole to the extent that this is governed by exam design and exam papers.**

In its approach to MFL assessment overall, at both GCSE and A level, Ofqual should take action to restore confidence in fair exams and fair grading in MFL. This must encompass ensuring that exam boards set papers which are fit for purpose for all levels of attainment, and that they implement appropriate quality control measures to ensure that the level of linguistic challenge in the exam papers is appropriate to what can reasonably be achieved by UK learners who do not have a native or near-native speaker advantage, in the classroom time typically available for the subject.
- 7. Ensure that MFL A level exam papers are appropriate for producing 'a rich and rewarding qualification, with an appropriate level of cognitive challenge and suitable for progression to university study or to employment' (ALCAB Report, p. 2).**

The level of linguistic challenge defined by the exam papers must be such as to allow teachers to focus on teaching content and fundamental language skills rather than on instilling exam strategies, in order to ensure that the reformed qualifications in MFL can serve the purpose for which they were designed.

Katrin Kohl
Professor of German
Faculty of Medieval and Modern Languages
University of Oxford
7 May 2019