

ANNEX C

GCSE Modern Foreign Languages Inter-Subject Comparability and the role of the Common European Framework of Reference for Languages (CEFR)

Response to a request for views, dated 19 March 2019

I am grateful for the opportunity to comment on Ofqual's investigation into GCSE grading standards in French, German and Spanish, and specifically on the evidence that should be included and its weighting. This investigation is greatly welcomed by the subject community in secondary and tertiary education.

Ofqual's concurrent study aiming to 'link several grades in GCSE French, Spanish and German to the CEFR to inform further discussions about performance and grading standard in GCSE MFL'¹ holds significant risks, and linking or benchmarking would in my view be premature in the present context. It should be considered only once confidence has been established in the reformed qualifications, and in particular the appropriateness of linguistic challenge (see Annex B for A level). Robust evidence must be in place that grading standards in GCSE modern foreign languages are appropriate, and suitable for learners without a native or near-native speaker advantage in the language being assessed, before any meaningful link can be made to CEFR levels.

Moreover, it will be vital to avoid assumptions about equivalence between GCSE grades and CEFR levels that result in the bar being set too high for all or parts of the exam cohorts and thereby exacerbate the existing problems concerning perceived and actual 'difficulty' of the exams.

Evidence to be included in any consideration of the case for an adjustment to grading standards in GCSE modern foreign languages

In reaching a decision, it will be important to ensure that application of the criteria, sources of the evidence, and use of the evidence are transparent and appropriate. Having analysed the Policy Decision on severe grading in the MFL A level, my view is that this was not consistently achieved in the earlier investigation, and I hope that my comments submitted as Annex A will be helpful for the forthcoming research.

Evidence concerning the 'difficulty' of MFL exams relevant for informing appropriateness of exam papers:

There is a long history of linguistic demands in MFL exams being pitched unrealistically high for the amount of classroom time typically available for UK learners. This was already highlighted in research published in 1998.² Ofqual will have received a response from Professor Suzanne Graham, Department of Education, University of Reading (a member of my AHRC-funded Creative Multilingualism research team) presenting research-based evidence that the decline in numbers of learners MFL GCSEs is in large part attributable to their perception that they will be unable to achieve in the exams set at that level. Their perception is founded on the actual excessive difficulty of the material they are confronted with in the GCSE exam (e.g. 12% of the vocabulary in a sample listening comprehension passage is from frequency levels outside the learners' likely level of competence). Inappropriate difficulty impacts negatively on reliability and fairness of grading. Relevant research on this issue should be consulted and should form an important source of evidence.

The reformed GCSE qualification in MFL was developed in a context of aiming 'to make the content more demanding in line with the Government's intention for the reformed GCSEs',³ but this should not automatically entail increased linguistic challenge, especially given the fact that classroom time has not been increased. Part of the aim of increasing the demand at GCSE with respect to content was to prepare students

¹ Invitation on 10 January 2019 to participate in an Ofqual CEFR linking study in March 2019.

² Milton, J., & Meara, P. (1998). Are the British really bad at learning foreign languages? *Language Learning Journal*, 18, 68-76 (<https://www.tandfonline.com/doi/abs/10.1080/09571739885200291>).

³ For this and the following quotations see <https://ofqual.blog.gov.uk/2019/02/14/evaluating-new-gcses-in-french-german-and-spanish/>.

more effectively for the A level syllabus. However, analysis of the 2018 A level papers for German suggests that these are far too difficult for UK learners at this level, and that there is evidently too little quality control concerning linguistic challenge (see Annex B). The A level papers should not therefore be used to inform the appropriate level of difficulty at GCSE until this anomaly has been effectively addressed.

It is a concern in this context that Ofqual's recent evaluation of the new GCSEs in French, German and Spanish highlighted the following responses from teachers:

- Ahead of the 2018 exams: 'There were concerns that changes such as the requirement for more questions to be in the target language, and listening material to be at 'near-native' speed, would make them overly difficult.'
- Following the 2018 exams: 'While teachers generally preferred the new GCSEs, there were concerns about some questions being too difficult, particularly in the listening and reading assessments.'

Teacher concerns about excessive difficulty of listening and reading assessments are supported by Suzanne Graham's findings with respect to the sample listening comprehension paper. They also chime with my findings concerning the 2018 A level papers in German (see Annex B). In the light of this, Ofqual should revisit its finding that 'the 2018 assessments were more difficult than in 2017' and its conclusion that this moved the papers 'to within an "ideal" range' owing to a better spread of marks. It will be important to obtain relevant evidence from teachers, and to conduct further research if they still consider the exams to be too difficult. The 2018 German A level papers suggest that the following should be investigated:

- The effect of questions being in the target language for learners at all levels of attainment, given that questions may contain words and phrasing not provided in the Specification. This is likely to significantly increase exam stress and classroom time spent on teaching exam strategy, and risks a disproportionate number of marks being unavailable to a candidate if a word is not correctly understood.
- Whether appropriate time is being allowed to complete the tasks.
- Whether listening comprehension recordings are appropriate in terms of accessibility of overall argument, thematic content, register, vocabulary, syntax, and whether questions are pitched appropriately.
- Consistent appropriateness of the vocabulary in terms of frequency. The reference to 'more demanding vocabulary in the reading and listening texts' in the evaluation blog post suggests that the demand may be excessive given the classroom time available to prepare students for the exams. The English exam boards do not currently appear to use word frequency statistics when determining the vocabulary used in exam texts and tasks. In the A level exams such an approach leads to excessive and sometimes random linguistic challenge.

The document 'GCSE Modern Foreign Languages. Consultation on Conditions and Guidance' states the following (Point 1.17): 'In new GCSEs, if vocabulary lists are provided by awarding organisations as a guide to teachers, assessment tasks should not be restricted to the vocabulary lists, as this could make assessments predictable. We have drafted requirements stating that at foundation tier, students should expect to deal with common/familiar words that are not on any vocabulary list, and at higher tier, candidates should be expected to deal with less familiar words, particularly in authentic sources.'⁴

It is not clear that restricting vocabulary at this level automatically makes assessment inappropriately predictable, and this approach should be revisited. Predictability of the content that will be assessed is not unusual in exams, and predictability of vocabulary can increase motivation to learn it thoroughly. The message given to examiners here is that they should *avoid* predictability, and potentially challenge candidates with authentic sources that go beyond what they can reasonably be expected to know. It should be considered normal at this level to adapt authentic sources as necessary especially in exam contexts to ensure that they don't overstretch candidates. The instruction also fails to give guidance on how to select 'common/familiar words' or 'less familiar words'.

It should be noted in this context that the Goethe Institut – which is highly respected for its successful language teaching and assessment – makes rigorous use of word frequency lists geared to CEFR levels in setting its exams.

⁴ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/387691/gcse-modern-foreign-languages-consultation-on-conditions-and-guidance.pdf , p. 13.

Evidence concerning the ‘difficulty’ of MFL exam papers and its impact on learners and society at large:

A key piece of supporting evidence which should inform the investigation is the finding in a recent BBC survey of secondary schools that 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.⁵ This is a critical piece of evidence from a methodologically impressive and high-impact survey: the BBC disseminated the survey to all mainstream secondary schools, and 2,048 out of some 4,000 responded.

The finding should be taken as evidence of the effects of the following four interlinking factors, all of which are within the sphere of Ofqual’s control:

- Severe grading in MFL (see Annex A, with respect to A level)
- The effect of the involvement of native speakers and near-native speakers in exam cohorts on fairness of grading and appropriate linguistic challenge in the exam papers for learners without such an advantage (see Annex A and B, with respect to A level)
- The excessive linguistic challenge posed by MFL exam papers (see Annex B, with respect to A level)
- The excessive requirement for teaching exam strategies necessitated by MFL exam papers that are too challenging with respect to number and range of tasks, and linguistic difficulty.

In interaction with the optionality of modern foreign languages at GCSE, these factors underlying the perceived and actual ‘difficulty’ of the subject have a significant negative impact on the following:

- Standard of knowledge and standard of language skills, especially deep understanding of the fundamentals of the language essential for progression to the next level
- Learner confidence in their ability in the subject
- Learner enjoyment of the subject and motivation to pursue study further
- Progression of the learner to an MFL A level as the next level of learning in the MFL pipeline
- Reputation of the subject as likely to contribute to a competitive university entrance profile (incl. negative impact on parental view of the subject as a suitable choice)
- Teacher confidence and job satisfaction, with knock-on effects on retention and recruitment
- Reputation of the subject in schools and provision, with knock-on effects on provision at tertiary level.

Evidence of the impact of the participation of native and near-native speakers on statistical measures:

Participation of native and near-native speakers of the language being assessed in MFL qualifications is highly valuable in strengthening the nation’s linguistic range and competence and it can also be an asset in the classroom. However, Ofqual’s failure so far to address this factor means that it is acting as a significant disincentive for learners who do not have such an advantage, especially at the top end of the attainment spectrum. It was found to have ‘a small, yet important’ impact on grading at A level⁶ and resulted in a one-off adjustment in 2017. Ofqual must urgently establish the impact of this factor at GCSE for all languages, and differentiate between languages. For example, it is likely to affect Mandarin significantly, which is now the fifth most popular GCSE language and enjoying targeted investment by the DfE.⁷ In general terms, and even in the absence of robust statistical evidence, it should be assumed that for learners without a native or near-native speaker advantage in the language being examined, GCSE grades in MFL are more severe than they appear to be on the basis of the current statistics, especially at the top end of the scale.

Evidence provided by the secondary sector concerning statistical measures:

I consider it a matter of high priority for Ofqual to take account of evidence from the secondary school sector

⁵ 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>; additional data courtesy of Clara Guibourg; editor Branwen Jeffreys (branwen.jeffreys@bbc.co.uk).

⁶ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/610050/Native_speakers_in_A_level_modern_foreign_languages.pdf.

⁷ <https://www.britishcouncil.org/education/schools/support-for-languages/partnerships-courses-resources/mandarin-excellence-programme>; <https://www.gov.uk/guidance/join-the-mandarin-excellence-programme>

in reaching a decision, especially evidence provided by ASCL, HMC, ISMLA and ALL. This dimension of evidence appears not to have informed the Policy Decision on severe grading at A level. Evidence provided to Ofqual over many years by ASCL and other bodies indicates that GCSE grades are approximately one half to one grade lower than they should be by comparison with other EBacc subjects. Given that Ofqual's statistical methods are designed in such a way that they replicate such an effect from one year to the next, a one-off adjustment is required to bring MFL grading standards into line with other subjects and ensure that MFL is not of above average difficulty.⁸ Among the evidence submitted to Ofqual is a joint letter by ASCL, HMC, ALL and ISMLA setting out the concerns about severe grading of MFL at GCSE and A level (6 July 2017). A survey on severe grading at GCSE MFL authored by ASCL in November 2017 and disseminated by ASCL, ALL, ISMLA and HMC generated over 2,600 responses from teachers in under a week, with over 90% reporting the negative perceptions pupils hold about the subject owing to severe grading.⁹

Evidence provided by stakeholders beyond the secondary school sector concerning the impact of severe grading on society:

Grading anomalies in MFL have become a matter of common knowledge far beyond schools and far beyond the immediate modern languages community over the past decade. They are now at the centre of an unprecedented level of concern about the languages crisis in the UK. The following evidence should be taken into account especially when it comes to assessing the impact on users of the qualification and the impact on society at large:

- Alerted by school teachers, the HE community in Modern Languages became aware of the critical impact of severe grading at GCSE and A level on the health of the subject in 2013. Meetings and workshops involving myself and others followed in the subsequent years, including meetings with Amanda Spielman and Cath Jadhav on 15 November 2013 and with Dame Glenys Stacey and Michelle Meadows on 15 February 2016. Leading Modern Languages subject associations in UK Higher Education and the HE members of the Ofqual subject group for MFL sent a letter on 5 December 2017 urging Ofqual to 'accept that there is now sufficient evidence to demonstrate the severity of grading in MFL, both at A level and at GCSE, and that the steps necessary to rectify this anomaly need to be implemented with urgency'.
- Annual *Language Trends* surveys published by the British Council.
- The All-Party Parliamentary Group on Modern Languages launched a framework proposal for a 'National Recovery Programme for Languages' in March 2019, including the following objective among the top four priorities among some 40 goals: 'In England, students should have a reasonable expectation that their GCSE grade in MFL will be similar to that in other subjects without any systematic variation.'¹⁰
- The Call for Action launched on 28 February 2019 by the British Academy, Royal Society, Academy of Medical Sciences and Royal Academy of Engineering¹¹ is a response to the fact that 'learning foreign languages at school has hit an 18-year low across the UK',¹² and it cites wide-ranging evidence of the social impact of this loss of vital language skills as well as the fact that the number of undergraduates in modern languages fell by 54% between 2008-9 and 2017-18; ten modern languages departments have closed in the last decade and a further nine have significantly downsized. A series of roundtables organised by the British Academy will be adding to the evidence in the course of the year. A roundtable I attended organised by Universities UK in this context specifically highlighted the urgent need to address severe grading in modern languages.

⁸ For the statistical criterion of 'average difficulty' see the Policy Decision document for grading in MFL A levels, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/757841/ISC_Decision_Document_20.11.18.pdf, pp. 8f.

⁹ '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>.

¹⁰ <https://www.britishcouncil.org/contact/press/mps-and-peers-urgent-call-national-recovery-programme-revolutionise-language-skills-uk>.

¹¹ <https://www.thebritishacademy.ac.uk/sites/default/files/Languages-UK-2019-academies-statement.pdf>.

¹² <https://www.thebritishacademy.ac.uk/news/national-academies-urge-government-develop-national-languages-strategy>.

- The Arts and Humanities Research Council has invested over £16 million in modern languages research, an unprecedented investment that is designed to strengthen a subject that is perceived as dramatically weakened especially because of the broken schools pipeline.

Weighting and interpretation of the evidence including that of stakeholder concerns and social impact:

Statistical evidence concerning the appropriateness of grading should normally have priority, and should ensure that the GCSE in MFL is not of above average difficulty for learners who do not have a native or near-native speaker advantage in the language being assessed. Evidence from schoolteachers concerning anomalies should be given high priority as an indicator of areas that require further investigation. Learner experience should similarly inform Ofqual's oversight of appropriateness with respect to the design and level of difficulty of exam papers. Evidence concerning social impact should be taken into account, and as for other evidence, it should be used appropriately, transparently and consistency.

Use of evidence regarding social impact in Ofqual's Policy Decision on severe grading at A level did not fully meet these criteria. Comments quoted by Ofqual in the press release announcing the Policy Decision on severe grading¹³ feature judgements which suggest that contrary to criteria (b) and (d) set out in the report, concern about take-up was deemed to have no legitimate place in the consideration of grading standards:

- 'Much of the concern about subject difficulty appears to be about the declining take-up of certain subjects. In my view lowering the grade boundaries is not the right way to make subjects more attractive to potential students'
- 'Dumbing down A level [...] grades [...] could lead to students being taken onto degree courses for which they are not equipped'.

The first comment suggests that Ofqual may have selectively sidelined the views of HE participants from Modern Languages in their research study, on the assumption that their academic judgement concerning standards was clouded by concerns relating to take-up. The argument echoes an interpretation of HE views in the Policy Decision report with respect to grading in French in the light of criterion (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'):

- 'There was strong support from participants for an adjustment to grading standards [...] – although with some indication when the reasons for this were discussed that this was a response to declining university entries, *rather than* because the demands of the current grading standard were too high' (p. 13, italics added).

This highly problematic interpretation of HE participants' contributions suggests that notwithstanding Ofqual's explicit criteria relating to the relevance of impact on society at large, concern about declining take-up is not judged to be a valid criterion when considering whether subject difficulty or severe grading should be addressed. In effect, HE concerns about take-up, including those expressed by HE participants in Ofqual's study, are evidently deemed here to *undermine* the case for lowering grade boundaries, and conversely used to support the case for making no adjustment to grade boundaries notwithstanding evidence of 'strong support' for such an adjustment. Ofqual should be in no doubt that the MFL community across sectors and subject associations is united in considering an adjustment of grade boundaries at both A level and GCSE to be *both* appropriate in itself *and* key for addressing attrition.

Reference to 'dumbing down' suggests lack of appreciation concerning the full range of negative impacts exerted by severe grading and excessive linguistic difficulty of exam papers on a subject that is optional at GCSE (see Annex A and B), and takes no account of the impact on preparation for further study. Lack of scope for thorough teaching of fundamentals of the language, and systematic undermining of learner confidence in their language skills, impedes effective preparation for a demanding university Modern Languages course.

If its findings concerning grading at GCSE in MFL are to command confidence from the HE sector, Ofqual will need to ensure that the relative weight of criteria is transparent, that they are applied consistently, and that one criterion is not used selectively and inappropriately to sideline another.

¹³ <https://www.gov.uk/government/news/reviewing-inter-subject-comparability>

Proposal to link grades in GCSE French, German and Spanish to CEFR

Ofqual's proposed study 'in which we would like to link several grades in GCSE French, Spanish and German to CEFR to inform further discussions about performance and grading standard in GCSE MFL'¹⁴ holds significant risks at the present time. The ALCAB panel for MFL considered mapping the reformed AS and A levels onto the CEFR but concluded that this should be undertaken only 'in the light of the introduction of the reformed GCSE and GCE'¹⁵ and taking account also of the Key Stage 2 reform. Linking would be premature until confidence has been established in reformed GCSE and GCE assessment. There is otherwise a very high risk of setting the bar too high for UK learners and exacerbating the languages crisis, for the following reasons:

- CEFR relies on a framework of quantified guided learning hours, but use of the framework must take account of the following: conditions and timespan in which learning is undertaken; prior language learning experience; other factors affecting rate of progress. There are currently too many variables across language teaching in the primary and secondary system to make mapping in terms of teaching hours reliable. Defining levels aspirationally rather than realistically with respect to what is actually being provided in schools would exacerbate the current tendency to set the bar too high (see Annex B).
- UK learners and language learning progress in UK school contexts differ significantly from those in other EU countries. Learners in non-Anglophone contexts gain effective experience of learning a foreign language through learning English, in which national education systems invest systematically, and which is very well supported by a universal perception of educational value and careers value, by the prominent role of 'global English' in digital media, and by a positive reputation English enjoys among young people. Any further language provision and learning can therefore build on a solid foundation of knowing how to learn a foreign language. Learners of a foreign language in UK schools enjoy none of these advantages when learning their first foreign language, a factor that very significantly slows down progress, especially given that teaching time is typically no more than two or two and a half hours per week,¹⁶ access to native speakers is often very limited, and many students have no opportunity to visit a country where the language is spoken.
- The considerable inadequacies of the 2018 A level exam papers in German with respect to consistent and appropriate linguistic challenge (see ANNEX B) demonstrate that the exam boards are currently insufficiently attuned to the level that is reasonable and appropriate for UK MFL learners, and do not have adequate control mechanisms in place to ensure that linguistic expectations are realistic. Whether this differs between languages needs to be investigated. Until this problem has been fully addressed, it would be inappropriate to compound the difficulties with benchmarking against an external system, especially since this was developed for EU learners rather than specifically learners in UK schools.
- Ofqual has so far undertaken no research on the impact of native and near-native speakers on GCSE performance in MFL exams. This may be impacting significantly on what is deemed by awarding bodies to be realistic for exam texts and tasks (see Annex B). This factor needs to be systematically understood and properly accounted for, with understanding of variation across languages, before any linking is undertaken. It is essential to ensure that any linking, mapping or benchmarking to CEFR takes candidates as its reference who have no native or non-native speaker advantage in the language being assessed.
- Any linking to the CEFR would need to avoid assumptions concerning alignment of UK language competence with that achieved at a similar age in other EU countries, and could hold the danger of unrealistic assessment of actual levels of competence in the current context. This would significantly impede the establishment of best practice in MFL teaching and assessment in the current UK context.

¹⁴ Invitation to participate in an Ofqual CEFR linking study in March 2019 (10 January 2019).

¹⁵ <https://alevelcontent.files.wordpress.com/2014/07/alcab-report-of-panel-on-modern-foreign-and-classical-languages-july-2014.pdf>, p. 3.

¹⁶ See annual *Language Trends* surveys and the recommendation of 'two to (ideally) three hours per week' in <https://www.tscouncil.org.uk/wp-content/uploads/2016/12/MFL-Pedagogy-Review-Report-2.pdf>, p. 3.

- It may be helpful to give a view in very general terms on the levels I would expect to be relevant with respect to the GCSE. I would expect the high grades to be no higher than A2 and the low grades in the region of A1. It should *not* be expected that B1 would normally be attainable for a candidate without a native or near-native speaker advantage in the language being assessed. Higher attainment would require very significantly more classroom time than is currently normal in UK schools. It may in this context be helpful to say that candidates accepted for a modern languages course at a university such as Oxford on the basis of having attained an A or A* at A level will typically be performing around level B1.

Recommendations:

To be appropriate, GCSE grading in MFL should fulfil the following criteria:

1. A learner without a native or near-native speaker advantage in the target language should be able to expect a grade in MFL that is statistically no lower than average GCSE grades in EBacc subjects.
2. The impact of the participation of native and near-native speakers in the exam cohorts on the statistical measures should be understood by Ofqual and the exam boards and addressed in a way that ensures fairness for learners who do not have such an advantage.
3. Evidence from schoolteachers and learners should form part of the evidence used by Ofqual to ensure that the design of MFL GCSE exam papers and the level of linguistic challenge are appropriate, and that grading is in line with that for other EBacc subjects.
4. Appropriate quality control measures should be in place to ensure that the linguistic demands are appropriate in the light of the number of classroom hours likely to be available to the average UK learner without a native or near-native speaker advantage in that language. These should particularly establish an appropriate challenge in the following areas:
 - a. Listening comprehension recordings and reading comprehension texts should be designed to give candidates across the competence range fair access to marks.
 - b. Word frequency (an appropriate reference for word frequency should be established and consistently used to ensure that vocabulary is within an appropriate frequency range)
 - c. Simple syntax and tasks that are designed to reinforce and test a thorough command of fundamentals
5. Rubrics and instructions for tasks in the exam papers should either be provided comprehensively in a list in the Specification, or given in English as well as the target language in order to avoid a disproportionately negative impact of not knowing a single word, avoid excessive exam stress, and reduce teaching time spent on exam strategies.
6. Formal or informal linking to CEFR levels should only be used once proper quality control is in place with respect to exam design, level of linguistic challenge, and effect of native and near-native speaker participation on grading standards and exam difficulty. While consultation of CEFR may help to establish appropriate grading standards, any formal linking, mapping or benchmarking must only be undertaken once confidence in the appropriateness, robustness and reliability of GCSE grading are in place.

Katrin Kohl
 Professor of German
 Faculty of Medieval and Modern Languages
 University of Oxford
 30 April 2019