

TECHNICAL REPORT

Inter-subject comparability

The evidence pertaining to the claim of grading severity in A level French, German and Spanish and the impact of statistical alignment of standards on outcomes

Beth Black, Jonathan Clewes, Richard Garrett, Milja Curcin,
Qingping He

Contents

Executive Summary	5
1 Introduction	7
2 Evidence pertaining to the claim of grading severity	9
2.1 <i>Statistical evidence</i>	9
2.1.1 <i>Evidence generated from Rasch modelling and other conventional statistical techniques</i>	9
2.1.2 <i>Evidence from comparative progression analysis (CPA)</i>	13
2.1.3 <i>Interpretation of results from statistical analysis and implications</i>	15
2.1.4 <i>Limitations of Rasch modelling and other statistical techniques</i>	16
2.1.5 <i>Limitations of CPA</i>	17
2.1.6 <i>Potential causes of difference in statistical difficulty between subjects and link to grading standards</i>	22
2.2 Judgemental evidence	23
2.3 Evidence of the impact of grading severity on society	23
2.3.1 <i>Inter-year A level entry data</i>	25
2.3.2 <i>Second and third language taken together with A level French, German or Spanish</i>	26
2.3.3 <i>Language Trends 2016/2017 Survey Report</i>	27
2.3.4 <i>University uptake</i>	36
2.3.5 <i>Evidence of filtering (selection and self-selection)</i>	40
2.3.6 <i>A level entry gender breakdown</i>	42
2.3.7 <i>Teacher supply data</i>	44
2.3.8 <i>Modern foreign languages pedagogy review</i>	51
2.3.9 <i>Research into subject choice motivations</i>	52
2.4 Views of stakeholders	56
2.4.1 <i>Subject associations' views</i>	56
2.4.2 <i>Perspectives from higher education</i>	63
2.4.3 <i>Interaction with other issues</i>	66
2.4.4 <i>Proposed changes</i>	68
2.5 Awarding Organisations' views	70
2.5.1 <i>Awarders' views</i>	70
2.5.2 <i>Viewpoints of the Awarding Organisations</i>	72
2.6 Higher Education Representatives' Perception of Grade Standard Adjustment	77

3 Impact of statistical alignment of standards on outcomes	79
3.1 <i>Change in grade boundaries and grade distributions</i>	79
3.1.1 <i>Aligning to the mean of all subjects (based on Rasch analysis)</i>	80
3.1.2 <i>Aligning to A level Geography (based on Rasch analysis)</i>	81
3.1.3 <i>Alignment of standards based on CPA analysis</i>	82
3.1.4 <i>Impact of small changes in outcomes</i>	82
3.2 <i>Change in performance standards as judged by experts</i>	84
3.3 <i>Summary</i>	84
4. Decision	84
5. References	85

Executive Summary

The purpose of this report is to gather together evidence to allow Ofqual to make an informed, evidence-based decision about whether there is a need to make an adjustment to grading standards in A level French, German and Spanish. Evidence to investigate the claim that these subjects are severely graded compared with other A level subjects was collected from a range of sources, including:

- Statistical evidence
- Judgemental evidence
- Impact of grade severity on society
- The subject associations'¹ perceptions of grading severity and impact
- Awarding organisations' views
- Higher education representatives' perception of grade standard adjustment

In addition, the impact of statistical alignment of standards between A level French, German and Spanish and other subjects in terms of their grade distributions and performance standards is also discussed in this report.

Results from statistical analyses using a range of methods suggests that A level French, German and Spanish are generally more “statistically difficult” (although they do not feature amongst the *most* apparently severely graded subjects) than many of the other A level subjects both in terms of a given grade, and at the overall subject level, and have been for at least the past decade. Such evidence has been frequently interpreted as an indication of grading severity in these subjects. The report discusses the major issues with the statistical methods used for comparing standards between subjects, including the strong assumptions made about the data being analysed (e.g. the unidimensionality assumption that a single underlying latent trait is assessed by examinations in different subjects required by the Rasch model and other similar models) which are seldom met by real data. Many of the factors that can potentially influence examination performance are also discussed, including motivation of students and efficiency and effectiveness of teaching and learning. It is suggested that these factors can vary substantially between subjects and that most statistical methods fail to take them into account.

No work involving the use of judgemental methods to compare the standards of A level French, German and Spanish with those of other subjects is available.

A number of negative trends within these languages have been attributed to the impact of severe grading. These claims are examined through reference to wider contextual data such as A level entry and university acceptances, figures on teacher supply and recruitment, changes in the cohort profile of students sitting these subjects, and qualitative data surveying schools and pedagogy. Whilst these sources confirm an overall decline in entries to A level French and German, it is challenging to prove causation – particularly when comparisons are drawn with A level Spanish. Research into subject choices also suggests that other factors take precedence over difficulty when students select their A level options.

Stakeholders who have raised concerns about the comparability of these subjects to other A levels argue for urgent action over grading standards in French, German and

¹ Details of the subject associations involved are given on page 59.

Spanish, asserting that these subjects are now at crisis point. In their view the statistical analyses vindicate long-standing concerns about severe grading, and maintaining the status quo is threatening the viability of post-16 language study. Stakeholders feel that students are being dis-incentivised from studying languages due to the perception that they will perform better in other subjects, or having the opportunity to study removed from them altogether by schools scaling back languages provision in response to declining entries. They regard the situation as particularly critical because of the potential for the current negative trends in languages to be self-perpetuating.

These views are shared by some in higher education, who are concerned about the impact that a decreasing number of A level students progressing to university will have on teacher recruitment. They also argue that recent political developments will exacerbate issues around language uptake at a time when linguists will be particularly vital to the national interest. However, some of the issues that concern stakeholders about severe grading also encompass subject content, assessment design and the impact of native speakers – factors which have already been separately addressed through qualification reform or other changes.

The views of awarding panels responsible for setting and maintaining standards in these subjects have also been taken into account. There was no consensus amongst awarders that standards in these subjects were severe, with some arguing that certain judgemental grades were currently too lenient, and performance unsatisfactory. Awarders acknowledged that languages are perceived as ‘hard’, but generally attributed this either to the inherent cognitive demands of the subject, or other factors such as inadequate resourcing and shortcomings in the legacy GCSE. They also expressed concerns about the impact an adjustment to standards might have on the ability of universities to make valid admissions decisions.

The awarding organisations themselves have been given the opportunity to formally state their views. The exam boards note the limitations of some of the statistical models used to measure subject ‘difficulty’ as a basis for achieving inter-subject comparability, particularly the extent to which they are biased by the non-random nature of students’ subject choices. They also point out the current analyses were conducted on legacy qualifications, and that the relative ‘difficulty’ of these A levels may have changed as a result of the reforms (although the policy intention was that the difficulty would remain the same) and the decoupling of AS and A levels. Changes to content standards at GCSE were also felt to have the potential to impact on students’ perceptions of A level. The boards recommend further research be carried out on the impact of these changes before any adjustment to grading standards was considered, particularly in view of the potential risks to performance standards, progression and public confidence from such a change. They also identify a tension between the stated purpose of A levels to identify attainment in a particular subject, and their use as a ‘currency’ for progression to higher education in general. Overall the boards are of the view that performance standards within a subject are more important than comparative standards between subjects, and that these should be prioritised.

The report also considers research conducted by Curcin, Black and He (2018) into how changes to grading standards in these subjects might be perceived by higher education, and the potential impact on the way A level grades are used in admissions. Representatives from higher education, subject associations, and

Ofqual subject experts reviewed samples of A level student work at and below several grade boundaries in each subject, indicating how those below compared to those on current grade boundaries. The participants were subsequently engaged in panel discussion on grade standards issues. The research showed that, overall, support for an adjustment to standards in a subject was inversely related to apparent difficulty under statistical measures, being stronger in those subjects which were closer to 'average' A level difficulty. Support for an adjustment to standards expressed in discussion also frequently contradicted the judgements made by participants in the script review. The research indicated that universities were likely to find adjustments to grading standards in A level languages more acceptable than in the sciences. In each of these subjects, there was agreement that the range of acceptable adjustment would be relatively small.

Finally, the report demonstrates that aligning standards statistically between subjects can have substantial impact on performance standards as represented by subject grade boundaries and grade distributions. The quality of performance expected of students, for each grade in A level French, German, and to some extent for Spanish, would need to be lowered substantially if they were to be aligned with other A level subjects statistically.

1 Introduction

From time to time, questions are asked as to whether some GCSE or A level subjects are harder than others and, if so, whether a better alignment should be achieved.

There is already a significant body of research into inter-subject comparability, to which Ofqual has contributed [here](#) with a programme of work begun in 2015. Ofqual sought to start a debate about the concept of inter-subject comparability through the publication of six working papers; a number of historical research papers on the topic, including some by Ofqual's predecessor organisation the Qualifications and Curriculum Authority (QCA); and a survey of views on potential policy options.

In response to the consultation on potential policy options, Ofqual resolved not to revise grading standards in all GCSE and A level subjects to achieve inter-subject comparability on a statistical basis. However, three A level sciences and three A level languages were identified as possible exceptional cases where we might take action to address concerns about the relative difficulty of these subjects – felt by some stakeholders to be the result of more severe grading.

To determine if this was the case, we recognised that it would be necessary to devise a basket of evidence upon which to make a decision about adjusting grading standards in a specific subject. This evidence would need to include not only the statistical data which initially prompted our consultation, but also the views of stakeholders (the subject associations and higher education selectors representing those subjects, and the exam boards and their awarders) as well as contextual data

which might contribute to our understanding of other factors that could have an impact on perceived difficulty.

These latter two sources of evidence were felt to be particularly important because it was noted that the comparisons between subjects on which the statistical data is based rest on conceptions of attainment-related linking constructs (such as 'general intelligence' or 'general academic aptitude') which some educationalists reject. Certainly, the plausibility and relevance of the evidence of 'difficulty' produced by these linking constructs would seem to diminish the less similar the subject that it is used to draw comparisons between (for instance, comparisons between physics and maths would seem to be more valid than those between physics and music). In light of this, it was recognised that whilst statistical evidence would be a key component of any basket of evidence, it must also be treated with caution. This report also explores some of the limitations of this statistical evidence in detail.

Nonetheless, statistical measures of subject difficulty are a source of evidence which, when considered alongside evidence of possible negative impacts within a subject, may contribute to a compelling case to adjust grading standards. For convenience, when discussing the evidence produced by these statistical measures we use the terms 'severe' and 'lenient', but it should be borne in mind that they are used in reference to the apparent difficulty of these subjects under Rasch and Comparative Progression Analysis only.

In considering this issue, there are different aspects of our statutory objectives and duties that we need to balance, reflecting that they are in tension with one another. We are required by legislation² to ensure regulated qualifications represent a "consistent level of attainment (including over time) between comparable regulated qualifications". If we were to adjust grade standards, we would need to consider that the issues in these subjects were sufficient for us to prioritise the need to achieve comparable standards in different but comparable A level subjects over the need to maintain standards in a subject over time. We must also reflect on the impact of any potential action on our objective to secure public confidence, and our duty to have regard to the views and needs of stakeholders who are 'users' of our qualifications, such as employers and universities.

We have since carried out a substantial programme of work to engage with stakeholders and seek their input in determining the basket of evidence for considering a possible adjustment to grading standards in A level French, German and Spanish, as well as revisiting the various statistical measures of subject difficulty.

The evidence we have assembled within each subject area includes statistical measures of subject difficulty such as Rasch and Comparative Progression Analysis;

² The 2009 Apprenticeships, Skills, Children and Learning Act.

contextual data on issues such as teacher supply and quality; and consideration of the views of stakeholders, including subject associations and higher education but also the exam boards and awarders. We have also considered evidence such as data on changes in A level entries and university applications over time; analyses of potential changes in the ability range and gender profile of the cohorts taking particular subjects; and research into motivations behind students' subject choices. Following advice from Ofqual's Standards Advisory Group, we have also incorporated into our evidence base information about recent grade boundaries in these subjects.

We have also conducted a research study into views on the impact of possible grading standard adjustment amongst higher education, to determine what the impact might be on the utility of these qualifications for university admissions were we to make any change. A summary of the findings is presented in this report.

2 Evidence pertaining to the claim of grading severity

This section provides a brief account of the evidence collected from a range of sources that has been used to support/challenge the claim that A level French, German and Spanish are severely graded compared with other A level subjects. This evidence can be broadly classified into the following categories:

- Statistical evidence
- Judgemental evidence
- Impact of grade severity on society
- The learned societies' perceptions of grading severity and impact
- Awarding organisations' views
- Higher education representatives' perception of grade standard adjustment

2.1 Statistical evidence

Statistical evidence can be broadly grouped into two categories, one associated with the conventional statistical techniques (including the Rasch modelling approach), one associated with the comparative progression analysis (CPA) approach.

2.1.1 Evidence generated from Rasch modelling and other conventional statistical techniques

Statistical methods used to study inter-subject comparability generally involve examining the relationships of grade outcomes between different subjects taken by the same students or the relationships between subject grade outcomes and

external variables that can potentially influence students' performance on the exams. Coe et al. (2008) provided a comprehensive review of the various statistical methods that have been used to investigate comparability. These include subject pairs analysis, common examinee linear models (including Kelly's method), latent trait models (e.g. Rasch models), reference tests and value-added models (including multilevel modelling) (see also Lockyer and Newton, 2015). We provide a brief explanation for each of these methods here.

The subject pairs analysis (SPA) approach looks at the average of the differences between the grades achieved in two subjects taken by the same group of candidates. When a large number of subjects are involved, the difficulty of a specific subject can be calculated as the simple mean or weighted mean of the averages of differences of all possible subject pairs. This is then compared with the difficulties of the other subjects calculated in the same fashion.

Common examinee linear models (including Kelly's method) derive the relative difficulties of different subjects from a matrix of examination by candidate results and involve finding the solution of a set of simultaneous linear equations.

In the case of using latent trait models such as the Rasch models, each examination is viewed as a polytomous item (characterised by a set of item difficulty parameters) in a test, and the grade or performance level assigned to an individual person (characterised by an ability parameter) on an exam are treated as scores on an item which represent ordered response categories. All exams contained in the analysis form a test. A mathematical function is used to describe the probability of a person (with a certain level of ability) succeeding on an item (with a certain level of difficulty). The difficulty measures of the items and ability measures of the persons can then be estimated using a range of approaches such as the conditional maximum likelihood (CML) estimation approach.

The reference test approach examines the relationship between subject grade outcomes and reference test scores (normally through regression analysis). Any difference in the relationship between subjects would suggest difference in difficulty. The value-added analysis approach (including multilevel modelling) represents an extension of the reference test approach. In this approach, the regression model can include a range of explanatory variables (such as candidate's prior attainment, gender, socioeconomic status, type of school attended, and many others) that can potentially influence examination performance.

For most of the statistical techniques, for the results to be interpreted appropriately, it is explicitly or implicitly assumed that the examinations in different subjects that are analysed together define a shared common construct or latent trait (such as "general academic ability/aptitude", the unidimensionality assumption) which is closely related to the constructs being measured by the individual examinations. The difficulty of a

subject is normally defined as the amount of the common trait required to achieve a specific level of performance on the exam. Difference in difficulty between exams in different subjects is assumed to reflect differences in “standards” which are related to this latent trait. It is to be noted that the latent trait is inferred from the analysis of the examinations included in the analysis.

Analysis of A level subjects using the Rasch model over the past decade or so suggested that Modern Foreign Languages (MFLs), including French, German and Spanish, are generally harder than the many of the subjects analysed (see Coe et al., 2008; He and Stockford, 2015). As an example, Figure 1 compares the overall difficulty and the difficulties of individual grades between the A level subjects from the 2017 exam series, with the subjects ordered by the overall subject difficulty. There is substantial variability in difficulty at individual grades between the subjects.

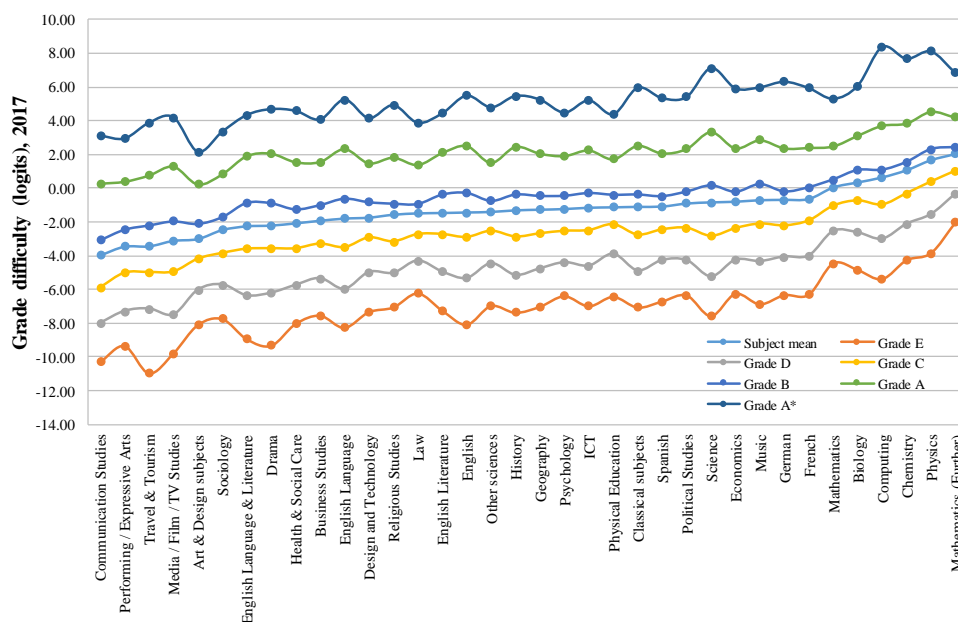


Figure 1 Comparison of the overall subject difficulty and the difficulties at individual grades for the A level examinations administered in 2017.

Table 1 below shows the overall relative difficulties and difficulties at individual grades for French, German, Spanish and geography in 2013, 2016 and 2017 respectively. Geography was chosen as a reference subject here because it is also a facilitating subject and of medium difficulty in most analyses. It is to be noted that, direct comparison of the values in Table 1 *between years* is not appropriate as the data for each exam series were analysed using the Rasch model separately. French, German and Spanish were substantially harder overall and at individual grades than Geography which was of average difficulty of all subjects contained in the analyses (which were set to zero) in 2013 and 2016. At A*, French and German were over half of a grade more difficult than Geography in 2013. However, the differences in difficulty became much smaller in 2017. At A, French and German were about a third of a grade harder than Geography in 2013, but this is reduced to about a tenth of a

grade in 2017. French and German were also more difficult than Spanish. Spanish is only slightly harder than Geography at A* and C.

Table 1 Relative grade difficulties (relative to the mean of all subjects contained in the analysis) expressed in grade width for A level Geography, French, German and Spanish from 2013, 2016 and 2017, based on Rasch analysis.

Subject	Relative difficulty (grade width)											
	2013				2016				2017			
	Overall	C	A	A*	Overall	C	A	A*	Overall	C	A	A*
Geog.	-0.05	-0.02	-0.04	0.21	0.06	0.03	0.02	0.11	0.05	0.06	-0.01	0.02
French	0.44	0.38	0.36	0.80	0.42	0.44	0.27	0.57	0.32	0.40	0.12	0.27
German	0.44	0.45	0.29	0.82	0.33	0.36	0.25	0.63	0.31	0.27	0.11	0.41
Spanish	0.28	0.30	0.24	0.61	0.19	0.15	0.13	0.34	0.14	0.15	0.00	0.08

Table 2, which was adapted from Coe et al. (2008), shows the relative grade difficulties of Geography, French, German and Spanish in 2006 analysed using a range of statistical methods. Again, Geography, French, German and Spanish were found to be more difficult than Geography and most of the other subjects.

Table 2 Relative grade difficulties (relative to the mean of all subjects contained in the analysis) expressed in grade width for A level Geography, French, German and Spanish from 2006, estimated using different statistical methods.

Subject	Relative difficulty (grade width)						
	Rasch	SPA (unweighted)	SPA(weighted)	Kelly	Reference test	Value-added	Multilevel
Geography	-0.13	-0.08	-0.17	-0.20	-0.02	-0.04	-0.05
French	0.51	0.34	0.14	0.26	0.20	0.42	0.49
German	0.50	0.26	0.09	0.24	0.17	0.44	0.47
Spanish	0.27	0.14	0.01	0.06	-0.07	0.16	0.16

Coe et al. (2008) also reported the relative difficulties of A level subjects estimated using Kelly's method from 1994 to 2006, and Table 3 shows the values for Geography, French, German and Spanish from 2000 to 2006. Again French and German were found to be about half a grade harder than Geography. Spanish on the other hand was of average difficulty. The difficulties are relatively stable over time, particularly since 2002.

Table 3 Relative grade difficulty expressed in grade width for A level Geography, French, German and Spanish from 2000 to 2006 estimated using Kelly's method.

Subject	Relative difficulty (grade width)							
	2000	2001	2002	2003	2004	2005	2006	
Geography	-0.23	-0.21	-0.30	-0.33	-0.28	-0.23	-0.26	
French	0.32	0.41	0.16	0.20	0.18	0.21	0.18	
German	0.15	0.27	0.15	0.15	0.23	0.13	0.11	
Spanish	0.07	0.10	-0.03	0.03	-0.03	-0.04	0.04	

2.1.2 Evidence from comparative progression analysis (CPA)

Newton, He and Black (2017) discussed the comparative progression analysis (CPA) approach to inter-subject comparability. CPA concerns the 'progress' made by individual students from their GCSE grade to their A level grade in the same subject area. It considers the distribution of A level grades in a subject, for students who were awarded a particular grade in the same subject at GCSE, to determine whether progression patterns are the same, or different, across subject areas. To interpret outputs from these analyses at face value – in terms of the alignment, or misalignment, of grading standards, or difference in subject difficulty – it is necessary to assume that groups of candidates ought on average to make the same progress across subject areas, all other things being equal; and that all other things are in fact more or less equal (see Newton et al., 2017).

Figure 2 shows the A level grade distributions for students with grades B and A at GCSE in the same subject in 2010 and 2016 for 14 A level subjects. The patterns of the distributions are similar for the two years. As is clear from the graphs, students taking French, German and Spanish and the science subjects achieved lower A level outcomes than those taking the subjects in English and social sciences. That is, students who studied A level French, German and Spanish and the sciences made less progress from GCSE to A level than those who had the same grade at GCSE in the same subject but studied English or social science subjects. However, the difference in A level grade distribution between subjects for students achieving an A grade at GCSE is not as prominent as that for students achieving a B grade at GCSE. To provide some quantitative information about the difference in A level outcomes between the subjects, Table 4 shows the average A level grade of the students in the subject with grades B and A at GCSE in 2010, 2013 and 2016 respectively. Overall, students who studied French, German and Spanish achieved about half of a grade lower than those studying English and social sciences.

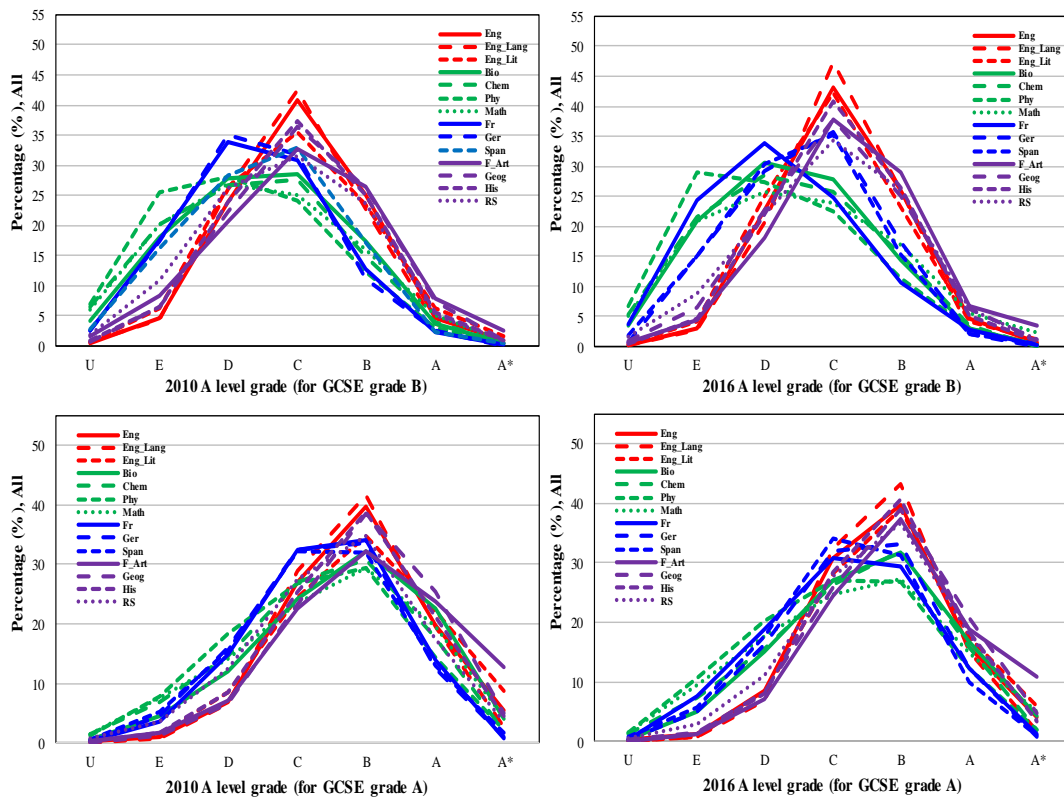


Figure 2 A level grade distributions for students who achieved grade B (top) and grade A (bottom) at GCSE in the same subject for 14 A level subjects in 2010 (left) and 2016 (right).

Table 4 Average A level grade in the subject for students receiving B and A at GCSE for 14 A level subjects (where A*=6, A=5 etc).

Subject	Subject average A level grade					
	Students with B at GCSE			Students with A at GCSE		
	2010	2013	2016	2010	2013	2016
English	3.02	2.96	3.07	3.86	3.77	3.72
English Lang.	2.95	2.98	3.06	3.80	3.70	3.69
English Lit	2.99	2.95	3.00	3.93	3.84	3.82
Biology	2.49	2.47	2.39	3.68	3.43	3.52
Chemistry	2.37	2.45	2.35	3.39	3.29	3.36
Physics	2.20	2.17	2.13	3.28	3.11	3.12
Mathematics	2.43	2.39	2.53	3.48	3.30	3.31
French	2.41	2.56	2.23	3.41	3.44	3.23
German	2.39	2.48	2.60	3.36	3.38	3.33
Spanish	2.54	2.49	2.55	3.35	3.36	3.23
Fine Art	3.08	3.24	3.24	4.07	4.10	3.98
Geography	3.06	3.09	3.03	3.95	3.93	3.83
History	2.99	3.07	3.06	3.82	3.80	3.77
RS	2.83	2.87	2.97	3.59	3.65	3.65

To look at how students progressed from GCSE to A level for the 14 subjects shown in Figure 2 across the range of GCSE grades, Figure 3 depicts the regression lines of A level grades against GCSE grades in the same subject. Students studying English and the social sciences generally made larger progress from GCSE to A

level than those studying French, German, Spanish, mathematics and the sciences, particularly for those with low to middle grades at GCSE. As the GCSE grade moves towards the higher grades, the difference in rate of progression between the subjects becomes smaller.

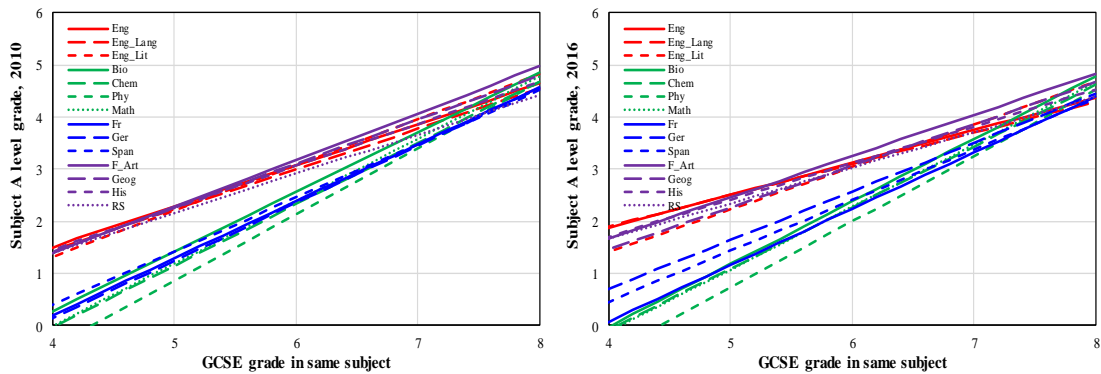


Figure 3 Regression of A level grade against GCSE grade in the same subject for the 14 A level subjects in 2010 (left) and 2016 (right).

2.1.3 Interpretation of results from statistical analysis and implications

In summary, in the case of Rasch modelling and other conventional statistical methods used to study inter-subject comparability, if we accept the assumptions involved, the differences in difficulty derived could be interpreted as differences in standards related to some traits assumed to be shared by the examinations in the different subjects. Such differences have frequently been used as evidence of inconsistency in grading standards (which are subject-specific) between A level subjects. Results from most of the analyses over the past decade or so suggested that A level MFLs, particularly French and German, are more difficult than many of the other subjects analysed.

In the case of CPA, again if we accept the assumptions made in the analysis, difference in the rate of relative progression from GCSE to A level would suggest inconsistency in grading standards at A level between the different subjects. Students studying MFLs showed substantially lower rate of progression from GCSE to A level than those studying the social sciences.

However, as is discussed below, there are limitations to the statistical techniques used to study inter-subject comparability.

2.1.4 Limitations of Rasch modelling and other statistical techniques

Coe et al. (2008) discussed the major issues with statistical methods used for studying inter-subject comparability. These, among others, include the strong assumptions made about the data being analysed (e.g. the unidimensionality assumption of the underlying latent trait shared by the examinees and assessed by the different examinations required by the Rasch model and other similar models) which are seldom met by real data, unrepresentativeness of samples used, missing data, imperfect data-model fit, sub-group effect, and different results from different statistical models for the same dataset (see also Lockyer and Newton, 2015).

In the case of using the Rasch model (and similar models) to study inter-subject comparability, the difficulty of a subject is defined as the amount of the trait common to students taking different subjects that is required to achieve a specific level of performance on the exam. Even if we assume that the A level data analysed meet the unidimensionality requirement of the Rasch model and fit the model, the interpretation of the latent trait specified in the Rasch model is not entirely clear. This is because the construct represented by the data implied from the Rasch analysis is inferred from the analysis of the set of examinations included in the analysis. This does not involve an actual measurement process in which the construct in relation to the purpose of the test to be measured must be specified and used to guide the development of the test (in this case defined by the complete range of subjects). The Rasch analysis is based on the relative frequencies of candidates receiving different grades in different examinations and it is difficult to interpret the latent trait inferred. Such a trait is likely to be influenced primarily by the subjects that are correlated well and have large entries. Therefore, to a certain degree, the extent of this shared common trait measured by the exams will likely vary between the subjects in relation to the traits which the individual examinations are designed to measure. With respect to the skills and knowledge assessed by the different examinations, although it is likely that some common skills will be assessed by examinations in different subjects, aspects of knowledge and understanding are generally subject specific and can vary considerably between subjects. Although the concept of “general academic ability” has been proposed to interpret the latent trait specified in the Rasch model (and other similar unidimensional statistical models), there has been limited research regarding the usefulness of the Rasch ability measures in relation to the specific uses the examinations are to be put to.

Because students normally take three or four A level subjects out of the 57 or so available, most statistical analyses involve large proportions of missing data which may not be assumed to be random. Using simulations, Bramley (2016) demonstrated that the existence of non-random missing data could produce biased estimates of subject difficulty using the subject pairs analysis approach. It is likely that similar bias in difficulty estimates generated using other statistical methods

would also exist. This may partly reflect the fact that differences in correlation between the subjects are different (or different subjects measure different constructs) and that missing data will influence correlations. One way to partially alleviate this problem might be to analyse semi-cognate or cognate subjects together to investigate relative subject difficulty.

2.1.5 Limitations of CPA

For CPA, to interpret the difference in progression from GCSE to A level between subjects as difference in grading standards, it is assumed that candidates ought on average to make similar progress across subject areas, all other things being equal; and that all other things are more or less equal. However, many factors that can influence exam performance can vary substantially between subjects. Such factors can include: level of subject demand, allocation of teaching time and other resources, motivation of students, efficiency and effectiveness of teaching and learning, uptake by various population subgroups and others, in addition to grading severity (or leniency) (also see Coe et al., 2008; Newton, 2012; Lockyer and Newton, 2015; Newton et al., 2017).

Selection of GCSE students for studying A levels

Table 5 shows the mean GCSE grades and standard deviations (SDs) of 13 GCSE subjects for all GCSE students and those who continued onto studying the 14 A level subjects discussed above in 2010 and 2016 (when calculating the mean GCSE grade, the GCSE letter grades were converted to numerical values with A*=8, A=7, B=6, ... and U=0). The average GCSE grades in the same subjects for those studying the 14 A level subjects are considerably higher than those of all GCSE students, with the differences nearly two grades for French, German, Spanish and about one grade for the sciences. The GCSE grade distributions for those studying the A level subjects were substantially narrower than those for all GCSE students. The average GCSE grades for those studying A level French, German, Spanish, the sciences and mathematics were well above 7 (GCSE grade A), substantially higher than the average GCSE grades of students studying English and the social sciences, suggesting difference in the selection from GCSE to A level between the subjects. The average GCSE scores of the students studying French, German and Spanish are also generally slightly higher than those of the students studying the sciences and mathematics. This may seem a little surprising from a learning progression perspective as the structure of knowledge and skills of the sciences and mathematics is more hierarchical than that of the languages. Differences in the selection from GCSE to A level could also contribute to the difference in the rate of progression from GCSE to A level between the subjects.

Table 5 Average GCSE grade in the GCSE subjects for all GCSE students and those who studied the 14 A level subjects (where GCSE A*=8, A=7, B=6 etc).

A level subject	Subject average GCSE grade							
	2010				2016			
	All students		Studying A levels		All students		Studying A levels	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
English	5.82	1.16	6.34	0.88	5.76	1.11	6.30	0.87
English Lang.	5.82	1.16	6.37	0.87	5.76	1.11	6.33	0.85
English Lit	5.86	1.21	6.72	0.92	5.87	1.22	6.80	0.85
Biology	6.53	1.12	7.10	0.80	6.33	1.16	7.11	0.76
Chemistry	6.58	1.11	7.34	0.75	6.32	1.18	7.33	0.69
Physics	6.54	1.13	7.33	0.75	6.33	1.17	7.32	0.71
Mathematics	5.69	1.34	7.27	0.72	5.77	1.34	7.46	0.65
French	5.62	1.46	7.45	0.71	5.54	1.47	7.53	0.67
German	5.63	1.37	7.33	0.78	5.60	1.38	7.37	0.77
Spanish	5.82	1.53	7.44	0.75	5.67	1.55	7.48	0.70
Fine Art	5.99	1.28	6.83	0.96	5.93	1.32	6.94	0.96
Geography	5.86	1.45	6.74	0.99	5.78	1.48	6.78	0.92
History	5.88	1.50	6.74	0.99	5.79	1.58	6.79	0.96
RS	6.09	1.43	6.99	0.96	5.98	1.48	6.88	0.92

Figure 4 further compares in more detail the distributions of GCSE grades in the same subjects for all GCSE students and those of the students who studied English, Physics, French and Geography at A level in 2010 and 2016 respectively. French and Physics are more selective than English and Geography. Further, French is also slightly more selective than Physics.

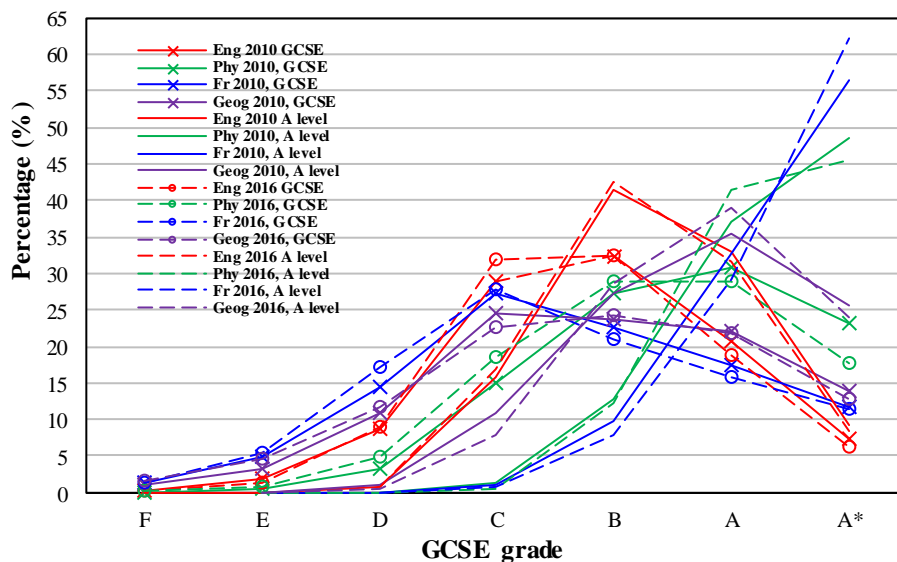


Figure 4 Distributions of GCSE grades for all GCSE students and those of the students who studied English, Physics, French and Geography at A level in 2010 and 2016 respectively.

Prior attainment

Figure 5 shows the distributions of mean A level scores of all students taking the 14 subjects in 2010 and 2016 and their mean GCSE scores obtained two years previously and the distributions for students with grades B and A at GCSE in the same subject. For all students, those who take the sciences, Mathematics and MFLs

generally achieved better outcomes at A level than those who took the English and subjects in social sciences. The pattern of the distribution of mean GCSE scores is similar to that of the distribution of mean A level scores, with students taking the sciences, Mathematics and MFLs having higher mean GCSE scores than those taking the other subjects. Students studying the sciences generally had considerably higher prior attainment than those studying English and the social sciences when they started their A level courses, with the average of the mean GCSE score nearly half a grade higher. Similarly, students studying the sciences achieved higher A level outcomes than those studying English and social sciences. The patterns of the distributions of mean GCSE score and mean A level grade across the subjects are similar for both years. For students who achieved grade B at GCSE, their mean GCSE scores are considerably lower than the mean GCSE scores of all students, particularly for those who studied the sciences and Mathematics. The mean GCSE scores for the students who studied French, German, Spanish, the sciences and mathematics are only slightly higher than those of the students who studied English and the social sciences, but their average A level grades are slightly lower. This may suggest that students with B at GCSE who took French, German, Spanish, the sciences and mathematics generally performed similarly across their A level subjects. For students who received an A at GCSE, those who studied the sciences and MFLs had mean GCSE scores similar to those of all students, but their mean A level grades were lower. The mean GCSE scores and A level grades for students who studied English and the social sciences were slightly higher than those of all students.

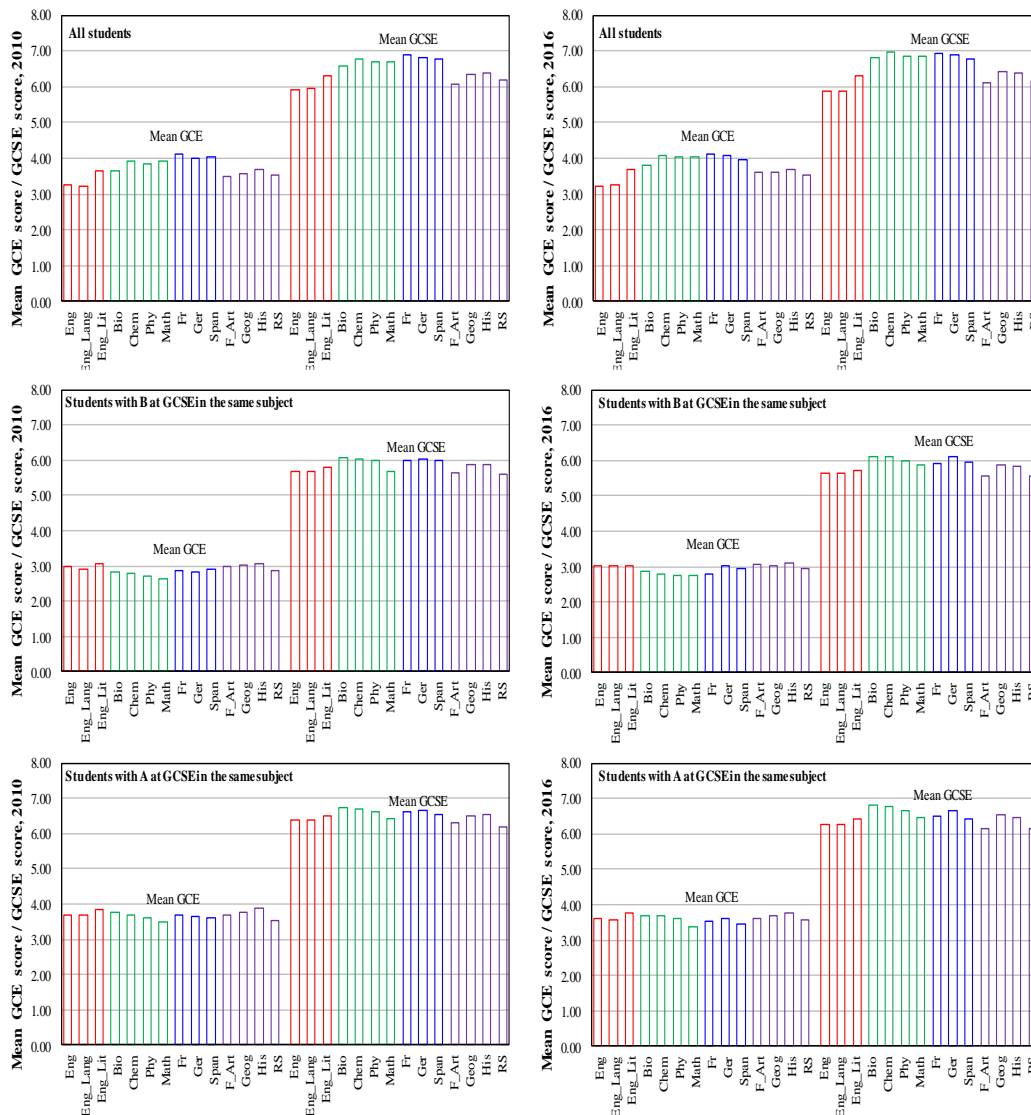


Figure 5

Distributions of mean A level scores of all students studying the 14 subjects in 2010 and 2016 and their mean GCSE scores (top), those who achieved grade B at GCSE (middle) and those who achieved grade A at GCSE (bottom).

To look at the relationship between A level outcomes and attainment at GCSE further, Figure 6 shows the mean GCSE score for all students at individual A level grades in 2010 and 2016. At individual A level grades, students taking French, German, Spanish, the science subjects and mathematics had substantially higher mean GCSE scores than students taking the other subjects, particularly at the lower grades where the difference was about half of a GCSE grade.

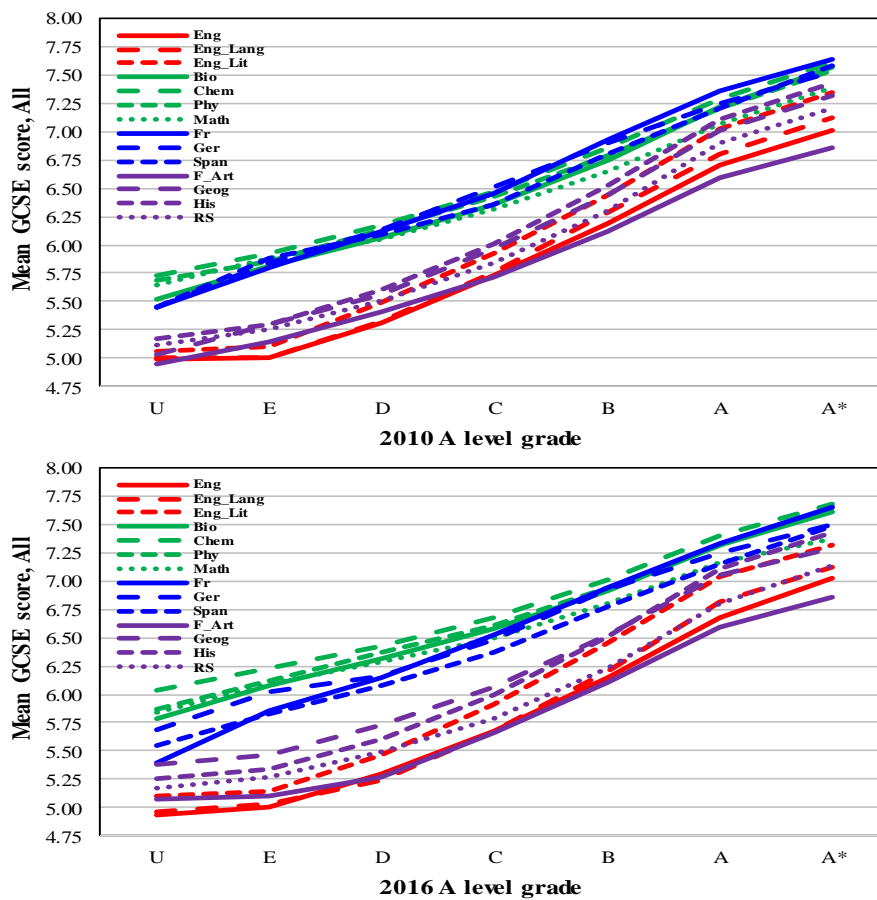


Figure 6 Average mean GCSE score at individual A level grade for all students studying the 14 subjects in 2010 and 2016.

The analysis presented above suggested that there are considerable differences in the attainment at GCSE between the 14 A level subjects. Such differences could produce different rates of progression from GCSE to A level even for students with the same grades at GCSE in the same subject.

Correlations of A level subject grades with GCSE grades in the same subject, mean GCSE score and average A level grade

To look at how A level grades relate to GCSE grades and other attainment measures, Table 6 below shows the correlations between A level scores and GCSE scores, mean GCSE scores, and mean A level scores for the 14 subjects from 2013. The correlations between A level and GCSE scores in the same subject were moderate, varying from 0.54 for Religious Studies to 0.62 for Geography. The correlations between A level scores and mean GCSE scores are slightly higher than the correlations with GCSE subject scores except for German, Spanish and Fine Art for which the correlations are slightly lower. This suggests that for most of the subjects, mean GCSE score rather than grades at GCSE in the same subject is a better predictor of performance at A level. As expected, the correlations between A level subject scores and mean A level scores are considerably higher than those

with GCSE subject scores or mean GCSE scores. Analysis conducted by Benton and Bramley (2017) indicated that difference in rate of progression from GCSE to A level based on CPA could also result from the difference in the level of correlation in the population between GCSE and A level. Further, their analysis shows that non-random choice of subjects could also produce difference in rate of progression from GCSE to A level between subjects.

Table 6 Correlations of A level subject score with GCSE score in the same subject, mean GCSE score and mean A level grade for the 14 subjects from 2013.

Subject	Correlation of A level subject score with		
	GCSE score for the same subject	Mean GCSE score	Mean A level score
English (Eng)	0.57	0.66	0.83
English Language (Eng_Lang)	0.57	0.66	0.84
English Literature (Eng_Lit)	0.57	0.73	0.87
Biology (Bio)	0.60	0.69	0.89
Chemistry (Chem)	0.57	0.64	0.91
Physics (Phy)	0.59	0.66	0.92
Mathematics (Math)	0.56	0.59	0.88
French (Fr)	0.59	0.62	0.84
German (Ger)	0.64	0.53	0.82
Spanish (Span)	0.56	0.53	0.82
Fine Art (F_Art)	0.61	0.59	0.83
Geography (Geog)	0.62	0.70	0.87
History (His)	0.59	0.69	0.88
Religious Studies (RS)	0.54	0.65	0.87

2.1.6 Potential causes of difference in statistical difficulty between subjects and link to grading standards

As demonstrated above, there are considerable differences in the difficulty indices at individual grade level and the overall subject level between A level subjects derived using different statistical methods. Such variability in difficulty has remained almost the same over the past decade or so. The differences in difficulty indices between subjects have frequently been interpreted as differences in grading standards. However, even if we accept the underlying assumptions of the statistical models and the level of model-data fit, these differences in difficulty between subjects can be caused by many factors. Coe (2008) and Coe et al. (2008) discussed a range of potential causes for such differences. These include differences in grading severity (or leniency), nature of the subject in terms of skills and knowledge to be learnt, level of subject demand, allocation of teaching time and other resources, motivation of students, efficiency and effectiveness of teaching and learning, and others (also see Newton, 2012; Lockyer and Newton, 2015). These factors are not considered in the conventional statistical methods and in the CPA approach.

It has also to be recognised that inter-subject comparability in A levels and other examinations has been a matter of debate for many decades. There has been no

consensus on how inter-subject comparability should be conceptualised, defined and measured (see Lockyer and Newton, 2015). Whether or not statistical difficulty indices should be linked to subject-specific grading standards of individual examinations has been the focus of such debate. Those who are against the use of statistical methods to compare different subjects argue that examinations such as A levels are graded based on standards that are subject-specific and that the shared knowledge and skills assessed by different examinations (the unidimensional assumption which is made implicitly or explicitly by most statistical models) is irrelevant, meaning that the between-subject comparison is of limited meaning. Further, they argue that there are many factors that can affect performance in exams which must be considered when comparing standards in different subjects. In contrast, those who support the use of statistical approaches argue that as long as there is a theoretical basis for the analysis and the interpretation of the results is justified, statistical comparisons would still be appropriate and meaningful. They further argue that subject standards may need to be statistically aligned when grades from different subjects are used for specific purposes, particularly when they are used interchangeably or as equivalent currencies in situations such as admissions to certain university courses and within school accountability measures.

2.2 Judgemental evidence

Judgemental methods used to investigate inter-subject comparability generally involve the use of experts to compare 1) the perceived level of performance of students on the exam (or quality of students' work) and 2) the perceived level of demand of the specification and other assessment materials between the subjects (see Bramley, 2011; Lockyer and Newton, 2015). Unlike statistical methods, judgemental methods look at the consistency of grading standards between subjects directly. The difficulty with judgemental methods is that it is difficult to establish cross-curricular performance criteria that can be used for inter-subject comparison and to find experts that are qualified to make such cross-subject comparisons (Jones et al., 2011; Lockyer and Newton, 2015). As a result, there has been very limited evidence of inter-subject comparability generated using the judgemental approach.

There has been very limited research involving the use of judgemental methods to investigate inter-subject comparability. None of the work we were able to locate involved any of the three A level MFLs.

2.3 Evidence of the impact of grading severity on society

Beyond the potential evidence of severe grading yielded by statistical measures of subject difficulty such as Rasch and Comparative Progression Analysis, languages stakeholders have identified several negative trends in A level French, German and Spanish which they assert are evidence of the impact of severe grading.

Primarily this evidence focuses on the declining uptake in modern foreign languages at A level and on undergraduate courses overall, but stakeholders also point to the

fall in the number of students studying two or more languages, as well factors such as issues with teacher supply; closure of university departments; and the potential for changes in the cohort studying languages as a result of filtering (whether students are selecting or being selected away from the subject) to distort the grading. Ostensibly some of these factors could be considered to be an alternative explanation of the severe grading suggested by statistical measures of subject difficulty. However, stakeholders argue that these are trends which result from the perception (confirmed by the statistical evidence) of severe grading, and that if they are not addressed they have the potential to lead to a 'feedback loop' which in turn will contribute to the 'difficulty' of these subjects, and put their continued presence within the curriculum in jeopardy.

The Joint Council for Qualifications publish a detailed report on entry trends for GCEs following each summer series, which includes data on how entries in a particular subject have changed over time. UCAS also publish data on applications and acceptances to specific subject areas. If present, we might expect to observe the effects claimed by stakeholders within this entry data – although if we accept the contentions of stakeholders about the historic nature of grading severity in A level French, German and Spanish, then negative impacts in recent entry data might not necessarily be compelling evidence, as the misalignment would have been carried forward since for a significant period of time (potentially since the inception of the qualification). Stakeholders would argue however that whilst severe grading within these subjects has been persistent, recent external developments have intensified the impact and led to further decline in entry.

Evidence of potential negative impacts from society might also be found in the longitudinal surveys of state and independent school modern foreign languages departments conducted by the British Council; official publications on teacher supply and recruitment; the recent review of modern foreign language teaching conducted by the Teaching Schools Council; and data on the changing cohort profile of candidates sitting A level French, German and Spanish.

However, where we do find evidence of potential negative impacts on society, it is not clear how strongly we could establish causation. There may be other factors present which could be producing similar effects, potentially either compounding those identified by stakeholders or having a direct impact on entries which may not be a result of severe grading. The view of stakeholders however is that in the absence of a comprehensive understanding of the causes of the decline in entry, the potential consequences of the loss of languages from the curriculum are severe enough that we should be employing the mechanisms within our gift (including adjustments to grading standards) to address the situation as best we can.

The subject associations have also argued that there are some impacts arising from severe grading which by their nature cannot be observed from entry data at all, regardless of the potential timescale over which this negative impact may have been in effect. This includes 'missing' candidates who would have studied the subject were it not perceived (rightly or wrongly) to be more difficult than others.

To some extent these missing entries might be inferred through comparison with other subjects which appear to be less 'difficult' under statistical measures. These are weak inferences however given the range of other factors which may be influencing subject choice, and require us to accept the argument (which, to some extent, conflicts with the findings of Cuff (2017)) that perceived difficulty overrides

students' enjoyment and interest, and the utility of the subject for further study or preferred career, when selecting options at A level.

2.3.1 Inter-year A level entry data

A level entries in French, German and Spanish are not significant enough to feature in the tables of intra-year entry data published as part of the Joint Qualification Council's GCE A level trend analysis following each summer series. However, JCQ does report specifically on year-on-year entry trends in A level modern foreign languages, reflecting the specific interest in their current status. Figure 7 shows the trend in A level languages entries from 2008 to 2017.

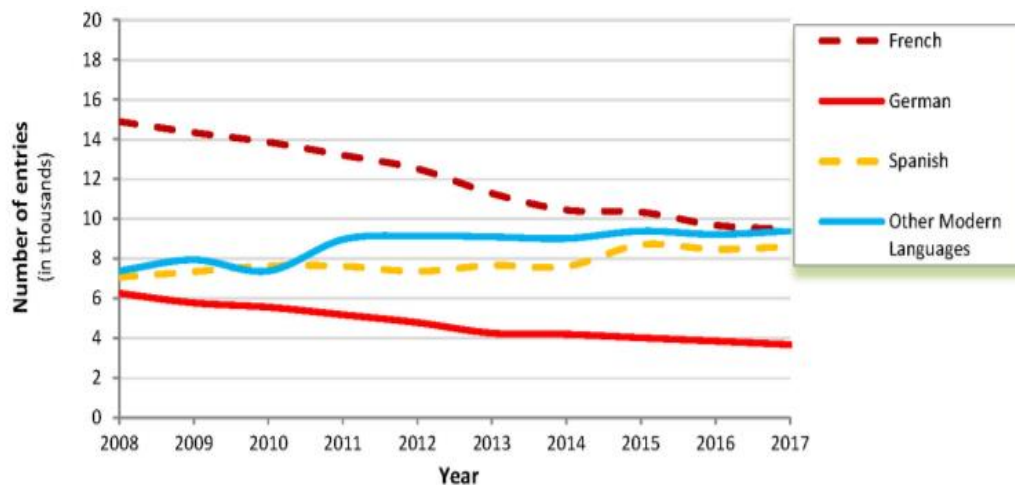


Figure 7 A Level language entries 2008 – 2017 (Joint Council for Qualifications, 2017)

Entries for A level French have undergone a significant decline since 2008, from approximately 15,000 candidates to just under 10,000. This decline is greater than that in A level German, although this may be reflective of the fact that French had a much higher entry to begin with.

The decline in entries in A level German has been less severe, but considered as a proportion of the total entry is similar to that in French. Moreover, entries in 2017 dipped below 4,000 students, and there is presumably a point at which the low numbers of students means that teaching the subject no longer becomes viable in many schools – not to mention assessing it, or offering the course at university. Notably, one exam board has already ceased to offer GCSE and A level languages, citing the decline in entry.

Entry trends in Spanish stand in marked contrast to those in French and German, having increased overall since 2008.

Entries for the broader JCQ reporting category of 'Other Modern Languages' have also increased overall since 2008, despite a minor dip in 2008.

It is important to bear in mind that these figures do not tell us about any changes in the type of students making entries for these subjects over the period shown. For instance, they do not tell us anything about the GCSE attainment range of A level language students in a particular year, and whether this is narrowing over time as

some stakeholders have asserted. Nor do they tell us about changes in the relative numbers of students within the cohorts from higher or lower attainment bands. Consideration of trends over time in the relative prior attainment profiles of students in these subjects is presented in a later section of this report.

Note that the JCQ figures given above do not take into account AS entries.

2.3.2 Second and third language taken together with A level French, German or Spanish

Table 7 shows the number of students taking A level French, German and Spanish between 2008 to 2017 in England and the percentages of students taking second and third additional languages using data extracted from the National Pupil Database (NPD). The number of students taking A level French and German have been declining from 2008 to 2017. The trend is different for Spanish which showed an increase in the entries from 2008 to 2015 and slight decrease in 2016 and 2017.

Table 7 Number of students taking A level French, German and Spanish from 2008 to 2017 and the percentages of students taking second and third additional languages (based on data extracted from the NPD).

Year	French			German			Spanish		
	N	2nd	3rd	N	2nd	3rd	N	2nd	3rd
2008	12641	25.12	1.87	5534	25.42	1.84	5733	30.86	2.55
2010	12401	23.23	1.69	5079	22.01	1.91	6579	27.47	2.43
2013	9870	22.74	1.98	3773	19.85	2.31	6505	24.75	2.38
2015	8987	24.87	1.64	3615	21.88	1.63	7600	22.78	1.50
2016	8172	24.02	1.90	3333	20.37	1.59	7196	21.98	1.65
2017	7801	24.78	1.41	2989	20.88	1.04	7044	22.49	1.35

Table 8 further shows the number of students who took other A level foreign languages in addition to French, German or Spanish in 2017. For students taking A level French, Spanish is the most popular other foreign language to take, followed by German, Latin, Italian and Russian. For students taking A level German, the most popular other language is French, followed by Spanish and Latin. For students taking Spanish, the most popular other language is French, followed by German, Italian, Latin and Portuguese.

Table 8 Number of students taking other A level foreign languages (ancient and modern) in addition to French, German or Spanish in 2017 (data from NPD).

	German	Spanish	Latin	Italian	Russian	Portuguese	Greek
French	450	1280	175	113	49	36	18
German		127	50	14	18	1	6
Spanish			92	110	47	78	13

2.3.3 Language Trends 2016/2017 Survey Report

Since 2002, the British Council has been conducting an annual research programme which looks at language teaching in English state and independent schools, and gathers teachers' views on the health of the subject. In their representations to us, a number of stakeholders have drawn our attention to the *Language Trends 2016/2017 Survey Report* compiled by Tinsley and Board (2017) as a source of evidence in itself, and also as a summary of some of the challenges currently facing language teaching. Two key stakeholders whose views are presented later in the report (the Association for Language Learning and the Independent Schools Modern Languages Association) contribute to this survey.

The report provides a useful perspective on stakeholder concerns because unlike much of the other data, in addition to an overall picture of entries in A level languages, it also provides a breakdown by school type. One contention of stakeholders is that severe grading of A level modern foreign languages is leading to an unequal student experience between the independent and state sectors, and this report allows us to examine these claims more closely.

Invitations to complete the *Language Trends 2016/2017 Survey* were sent out to 2970 state secondary schools and 655 independent secondary schools, with the survey itself being carried out between September and December 2016. The responses presented in the report came from 701 state secondary schools and 145 independent schools.

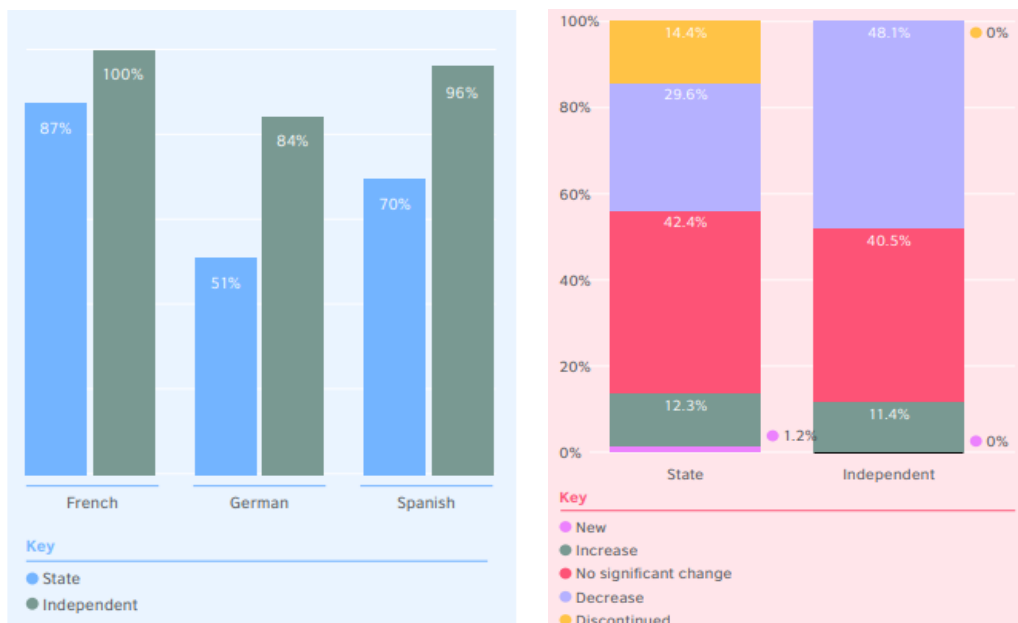
The annual nature of the *Language Trends 2016/2017 Survey* means that the majority of the figures it presents offer a snap-shot of a single year, rather than data which might allow us to observe ongoing trends within A level languages. Nevertheless, the contention from some stakeholders that the issues with severe grading in modern foreign languages are longstanding (potentially originating when the A level qualifications were first devised) means that although recent trends in languages may be more alarming to schools and universities, if we are looking for evidence of persistent grading severity then it is also worth considering the standing of the subject overall and in comparison to other subjects.

As well as data on language provision, uptake, and other aspects of delivery, the survey also gathers qualitative responses from teachers on each of the key aspects of languages teaching it reports on. Representative quotations from teachers are presented in the report, which are reproduced here where relevant.

As discussed elsewhere in this report, one concern raised by stakeholders is that the issues they feel arise from severe grading of A level French, German and Spanish are having a more detrimental impact on state schools rather than independent schools. The left graph in Figure 8 shows the relative standing of French, German and Spanish in state and independent schools according to the survey.

Perhaps unsurprisingly, language provision overall is greater in the independent sector than the state sector. The survey report indicates that this is likely to be due to financial pressures in maintained schools, with one respondent illustrating some of the consequences of these funding considerations: "French and German at key stage 5 are being taught jointly with the other comprehensive school in town."

Spanish is being phased out (despite its popularity and good results) due to funding cuts and a staff reduction process” (Tinsley and Board, 2017).



Figures 8 Post-16 provision for French, German and Spanish, independent and state schools with post 16 pupils (left), and changes in the overall numbers of pupils taking a language post-16, state and independent schools (right) (British Council, 2017).

Overall provision aside, the popularity of individual modern foreign languages is similar between state and independent schools, with French being offered most frequently in both sectors, followed by Spanish and then German. However the relative uptake of German as a proportion of the other languages is significantly lower in the state sector (51%) than the independent sector (84%). The gap between French and Spanish is also far slimmer in the independent sector than the state sector, with Spanish only 4% less popular than French. Notably, French was still offered at A level in nearly all of the independent schools included in the survey.

The survey findings also suggested that up to 12% of state schools may be offering no post-16 language provision in these subjects at all. This compares with just two out of the 127 independent schools who participated in the study who did not offer languages A levels. This suggests that the factors contributing to the decline entry in the two sectors are either different, or mitigated to an extent, in independent schools (where anecdotal evidence suggests that modern foreign language study is more likely to compulsory at GCSE than in state schools).

The right graph in Figure 8 shows the overall change reported in the survey of the number of students taking language A levels in state and independent schools. Whilst provision in independent schools may be higher overall, the figures indicate that a greater number of independent schools have experienced a decline in the numbers opting to study these subjects than in state schools. However, when the number of state schools who have ceased to offer languages altogether is taken into account, the numbers are much closer. This suggests that whilst independent and state schools may be experiencing a similar decline in A level language entries, independent schools are more likely (presumably through greater financial resources) to be able to continue to offer modern foreign languages classes with

fewer students. Figures otherwise seem very similar, with a fraction more state schools reporting an increase in A level language entry, and a very small number of state schools offering A level languages for the first time (or after a period of not doing so) – something not reported in the independent sector. For overall language entries then, the degree of fluctuation in entries seems in fact to be the same between independent and state schools. Where they differ is in the greater ability of independent schools to weather changing entry.

Another characteristic of the concerns raised by stakeholders is that they indicate not all modern foreign languages are suffering as badly as the result of alleged grading severity. A level German in particular is highlighted as a subject where the negative impact of inequitable grading has been particularly pronounced. This suggests either that there are differences in the degree of grading severity between A level French, German and Spanish, or that other factors are mitigating against the impact to a lesser or greater extent in these subjects. Figures 9 and 10 show the changes in the numbers of pupils taking French, German and Spanish respectively in state and independent schools in 2016.

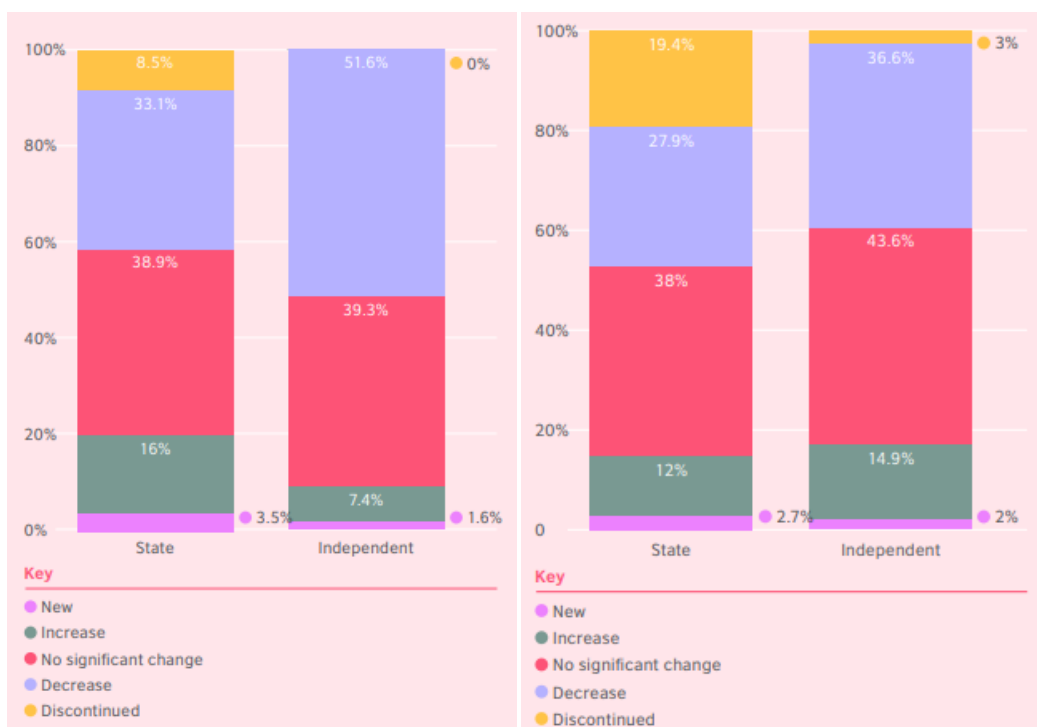


Figure 9 Changes in the numbers of pupils taking French post-16, state and independent schools, 2016 (left) and changes in the number of pupils taking German post-16, state and independent schools, 2016 (right) (British Council, 2017).

The greatest decline has been in A level French in both state and independent schools, with a decrease of 33.1% of A level uptake reported in the state sector and 51.6% in the independent sector. Though the decrease in the independent sector seems to be particularly severe, this is in part a result of the fact that French is the

most popular A level language and offered in nearly all independent schools – so any overall decrease in entry is more likely to occur within this subject. When the number of state schools ceasing to offer French entirely is also included, the figures are closer (41.6%). The number of schools experiencing no change is similar between the two sectors, although there is both greater reported growth and a greater number of schools offering the subject for the first time in state schools.

Responses by French teachers to the survey indicate that the subject is becoming less popular amongst pupils, and that this is leading schools to cease to offer it as small class sizes no longer become viable: “We have been told that we will not be able to continue with A level languages if we don’t recruit at least ten pupils.” The perception amongst teachers is that the reason for this decline in popularity is that the subject is considered to be harder, and that this has been the view for some time: “During the last five years, we have had not had sufficient pupil take up of French to run a French A level course, other than every two years... French is considered harder, and pupils do not want to run the risk of lower grades.”

The next most significant decline is in A level German, where the fall in entries is again ostensibly greater within the independent sector at 36.6% compared to 27.9% in state schools (perhaps reflecting the fact that this language is already much less popular at A level in state schools), but when combined with the number of schools which have ceased to offer the subject the decrease in the state sector is greater overall at 47.3%. The number of state schools reportedly ceasing to offer the subject this year is significant at 19.4%, but it should be noted that there are a small number of independent schools ceasing to offer German, at 3% - something which has not occurred in A level French. Given from this data that it appears the independent sector seems to be able to better weather decreasing entries, and that A level German provision is not significantly lower than French or Spanish, this may be indicative of other factors contributing to the lack of centres choosing to offer the subject beyond just declining entries. A greater number of independent schools than state schools reported stability in their entry for A level German, and very slightly more independent schools that their entry overall was increasing. Interestingly a similar number of schools (2.7% state, 2% independent) indicated that they would be offering A level German for the first time or after a hiatus, suggesting that there are some schools making an active choice to offer the subject despite the challenges identified elsewhere.

German teachers’ responses to the survey provide some possible explanation for this perhaps surprising increase in uptake in the face of an overall decline, primarily attributing increased entry to deliberate attempts to address shortcomings in teaching: “Better teaching at key stage 4 German has resulted in a higher uptake from boys in key stage 5.” This suggests that where entry is declining there may be other plausible causes beyond severe grading, and that where known issues with teaching quality in modern foreign languages are being addressed, individual schools are seeing increases in entry which buck national trends.

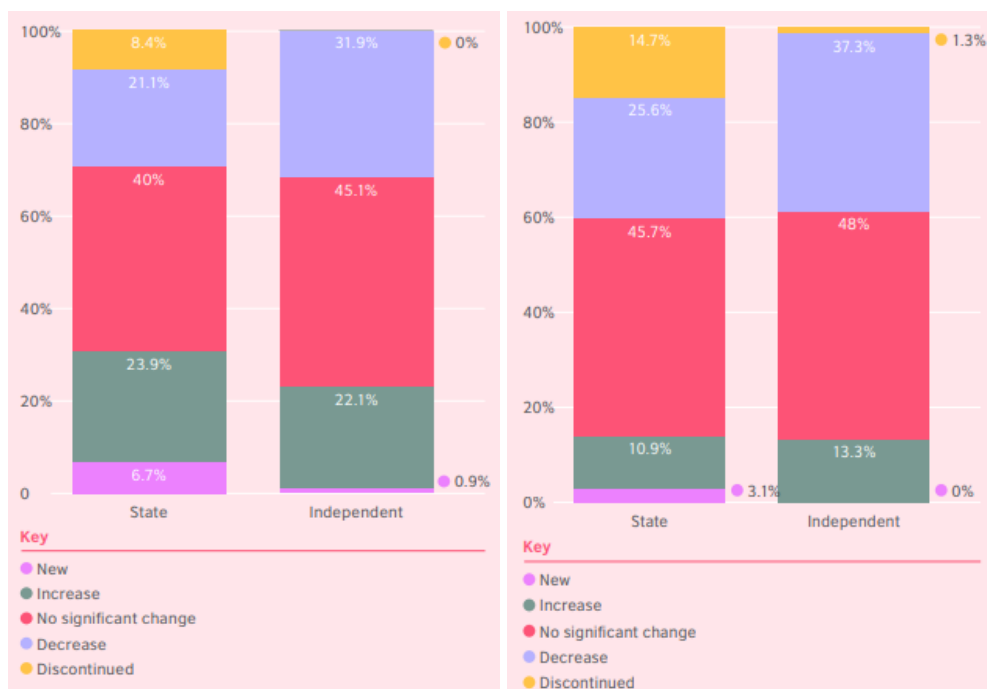


Figure 10 Changes in the numbers of pupils taking Spanish post-16, state and independent schools, 2016 (left), and proportions of schools reporting changes in the number of pupils taking more than one language post-16, state and independent schools, 2016 (right) (British Council, 2017).

Spanish is the only language where the proportions of schools reporting an increase in student numbers or offering the subject for the first time approach those of schools reporting an overall decrease (see the left graph in Figure 10). The independent sector has still reported an overall decrease in numbers taking A level Spanish (31.9%), but 22.1% of schools report an increase, and the majority (45.1%) that entries remain stable. In the state sector, the combined number of schools reporting an increase in entries or offering the subject for the first time (30.6%) just exceeds the combined number of schools reporting a decrease or ceasing to offer the subject (29.5%). Again, the majority report that their entry remains stable. These figures are interesting when considered against statistical measures of subject difficulty such as Rasch analysis. Under these measures, A level Spanish appears significantly more leniently graded than French or German. On the one hand, Spanish seems to be undergoing similar declines in entry in the independent sector to these other, apparently more severely graded languages – albeit in a less pronounced way. This could suggest that severity of grading is not the sole (or even primary) cause of this decline. On the other hand, the figures from the state sector and from A level entries overall, which show that Spanish A level entries are increasing when French and German are not, could be interpreted as evidence that it is the difference in grading severity between the subjects which is contributing to this situation – and that if French and German were as apparently leniently graded as Spanish, uptake in those subjects might also increase.

Spanish teachers responding to the survey were optimistic about the continued growth of the subject at A level, and some again attributed the cause of this to the quality of teaching at key stage 4: “In Spanish, numbers are now very high due to excellent GCSE results and new excellent teachers in the department who took the

pupils in Year 10 and 11.” Others however expressed concerns that the move to linear A levels would threaten this growth in the future, as students take fewer subjects: “Spanish was popular this year as a fourth AS subject... we are concerned that Spanish will decline with pupils only having the option of studying three subjects next year.” This is a factor which will be discussed in greater detail later on.

Another concern particular to language stakeholders with regard to the potential impact of severe grading is the decline in the numbers of students choosing to study two A level languages, which they regard as being a significant threat to the long-term viability of modern foreign languages in schools. Not only is this a potential explanation of the overall decline in figures (with the total number of students studying a language not necessarily changing significantly, but the number studying two dropping off markedly and consequently overall entries falling), but it has also been suggested as a potential exacerbating circumstance for issues with quality of teaching and other factors which would contribute to the ‘difficulty’ of the subject as experienced by learners, and which may offer a plausible alternative explanation to alleged severe grading. The right graph in Figure 10 shows the change in the number of post-16 students studying more than one language in state and independent schools.

Again the reported decline is greater in independent schools than state schools (37.3% as opposed to 25.6%), but this does not take into account that dual language study is no longer a possibility in 15% of state schools surveyed due to less popular subjects being discontinued. Where figures for declining entry and discontinued subjects are combined, the numbers are similar between the independent and state sector. Again, this is suggestive that the pressures in the two contexts are the same, but that independent schools are better able to withstand changes in entry. Similar numbers of state and independent schools reported that their numbers of dual linguists were unchanged, and again similar numbers that their entries were increasing. In the state sector 3.1% of schools reported ‘new’ dual linguists (i.e. had created an option which had not existed previously for students), which did not occur in the independent sector – but again this is likely to be the result of the fact that language provision in the sector has historically been greater.

Those teachers responding to the survey attributed the decline in dual linguists to subject difficulty, interacting with the move from four to three A levels as schools prepared to adopt the new linear qualifications:

“Double linguists are in decline as A level languages are perceived as more difficult to get top grades and so two languages can affect university choices.”

“Languages are seen as hard, with lower exam results compared to other subjects. There are no longer any dual linguists at key stage 4 wanting to continue to study at key stage 5.”

This decline in dual linguists is an additional nuance to bear in mind when considering the issue of inter-subject comparability, as it may be an additional consequence of severe grading which would be easy to overlook if we consider only the overall entries within individual subjects. Students may be making the choice to study only one language, balancing the potential impact on their results against their enjoyment of the subject or the potential utility for future careers, where they would previously have chosen two – but have deemed the risk to further study to be too great. This in turn may be having an impact on those progressing to higher education

to study modern foreign languages, where selecting institutions have historically required applicants to have two languages. This could be exacerbated (or an alternative explanation offered) by the increasing tendency to take three rather than four A levels post-16.

Some commentators have suggested that the entries in A level languages will begin to increase in the state sector in future years, as a result of increasing uptake at key stage 4 as more schools enter students for the EBacc. The *Language Trends 2016/2017 Survey Report* identifies 240 state schools where language entries at GCSE increased, and asked those schools whether this has led to signs of increased uptake at A level, presenting the figures alongside responses to the same question in the previous three years. This is shown in Figure 11. 13% of respondents indicated that they experienced a positive impact on Key stage 5 entries from the increased uptake at key stage 4, but the majority said they had not. This figure was down from 61% in the previous year, but in line with the responses in 2013 and 2014. Interestingly, the number of respondents who indicated that it was too early to say what the impact of increased uptake at key stage 4 would be was also up on previous years, potentially reflecting increased uncertainty about the impact of the move to linear A levels.

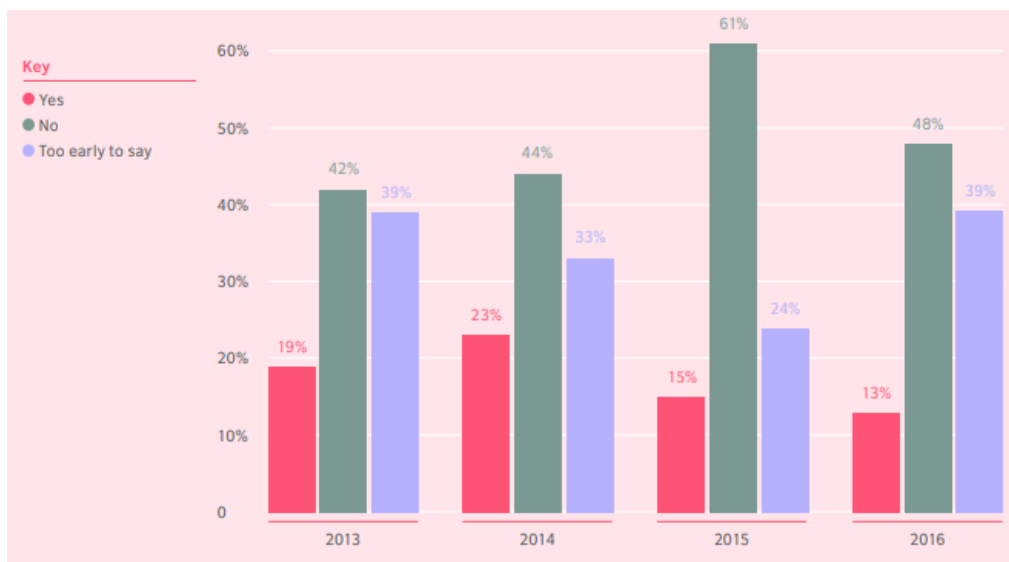


Figure 11 Whether increases in take up at key stage 4 have also improved take up post-16, state schools that have increased take up for GCSE, 2013-2016 (British Council, 2017).

In their responses to this issue, teachers again pointed to the perception that these subjects are harder than others:

“Many pupils with A/A* grades still refuse to opt for key stage 5 languages... always citing that languages are just too hard at A level.”

“We struggle to keep our best linguists... reducing options to three subjects, has affected us all – removing languages as the fourth option which they previously would have chosen as a subject they enjoyed, as opposed to a subject they needed.”

The clear argument here is that students are prioritising subject difficulty over enjoyment or utility when choosing their A level options, and that this has become

even more pronounced with the move to take fewer A levels. However, this is not supported by either the Ofqual research into subject choice or the DfE survey (this evidence is explored in greater detail in section 2.3.9).

Another frequently cited challenge to A level languages (and one which has been offered as a potential impact of grading severity, but could also be contributing to the apparent ‘difficulty’ of these subjects) is the problem in recruiting and retaining adequately qualified staff. There is an additional nuance here in that not all languages teachers, even ‘dual linguists’, may be able to teach a language to the same level, but are frequently asked to do so by schools when they experience issues with recruitment. The *Language Trends 2016/2017 Survey* asked schools about their situation regarding teacher supply and retention in languages. Figure 12 shows the responses for both state and independent schools for all languages.

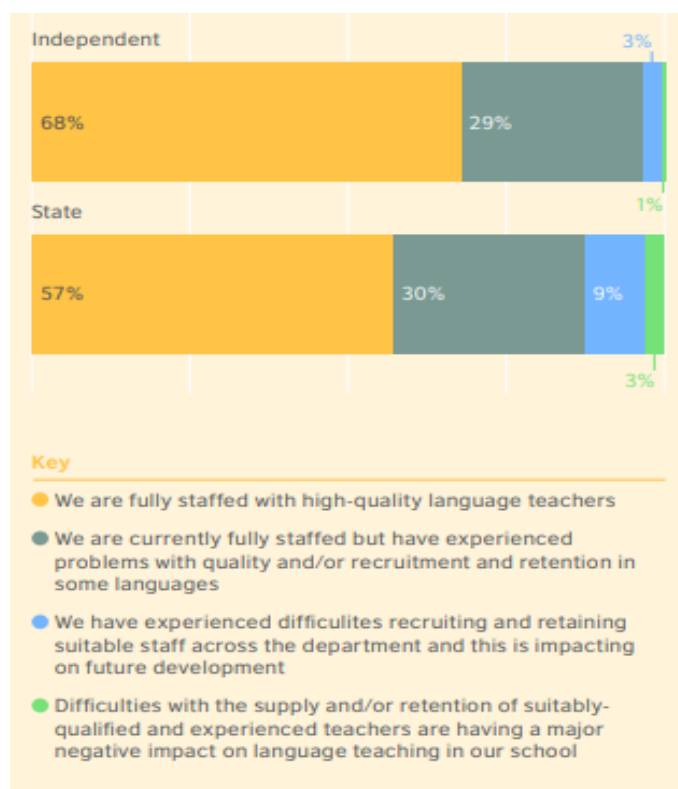


Figure 12 Teacher supply and retention, state and independent schools, 2016 (British Council, 2017).

68% of independent schools and 57% of state schools reported that they had no issues with securing high quality teaching staff. However, 29% of independent schools and 30% of state schools had experienced difficulties in some languages, and in 4% of independent schools and 12% of state schools issues with recruitment were either impacting future development or were having a major negative impact on teaching. This was supported by some of the responses to the survey: “We lost two teachers last year and it was very hard to recruit high-quality replacements. It is hard to find staff who can teach two languages up to GCSE and almost impossible to find staff who can do two languages to A level.” The responses of some teachers suggest that this may be contributing to the decline in A level entries for wholly practical, rather than academic, considerations: “We consistently had problems finding good German

teachers, which is partly why German has been dropped from the curriculum.” Others suggested that this was having an impact on pupil performance at specific grades, again offering a plausible alternative explanation for supposed grading severity: “Finding staff who can teach A* - A GCSE and A level has proven to be VERY difficult. We had gaps for two years.”

This is significant as, considered alongside the findings of the *Modern Foreign Languages Pedagogy Review* compiled by Bauckham (2016), the lack of adequately qualified teachers and systemic issues with approaches to language teaching may be a plausible alternative explanation for the issues present in A level French, German and Spanish. Stakeholders also point to the decline in entries for this subject as being a contributing factor to this shortage of adequately qualified teachers, or likely to make it worse if issues around uptake are not addressed. Certainly, some of the comments received in response to the survey suggest that this was the case: “... language sections at universities are closing down, so I am extremely worried about future recruitment.” One also raised an additional challenge in the form of attracting teachers in a shortage subject which is now in demand in business: “The dearth of language graduates willing to go into teaching is beginning to bite... there is seldom a spread of candidates from which to choose when vacancies come about.”

Finally, the *Language Trends 2016/2017 Survey Report* also looks at another challenge frequently cited by teachers – the general move from four to three A levels under the reformed specifications. This may offer some indication of likely future developments in modern foreign languages. The most relevant finding to the question of inter-subject comparability is the indication that 24% of independent schools and 15% of state schools are considering withdrawing or phasing out AS courses (Figure 13), with a likely overall impact on the number of pupils studying languages not just at A level, but at all. This in turn could have a detrimental impact on the continued viability of A level language departments (where AS levels are designed to be co-teachable with A level), leading to a further decline in entry for reasons other than severe grading.

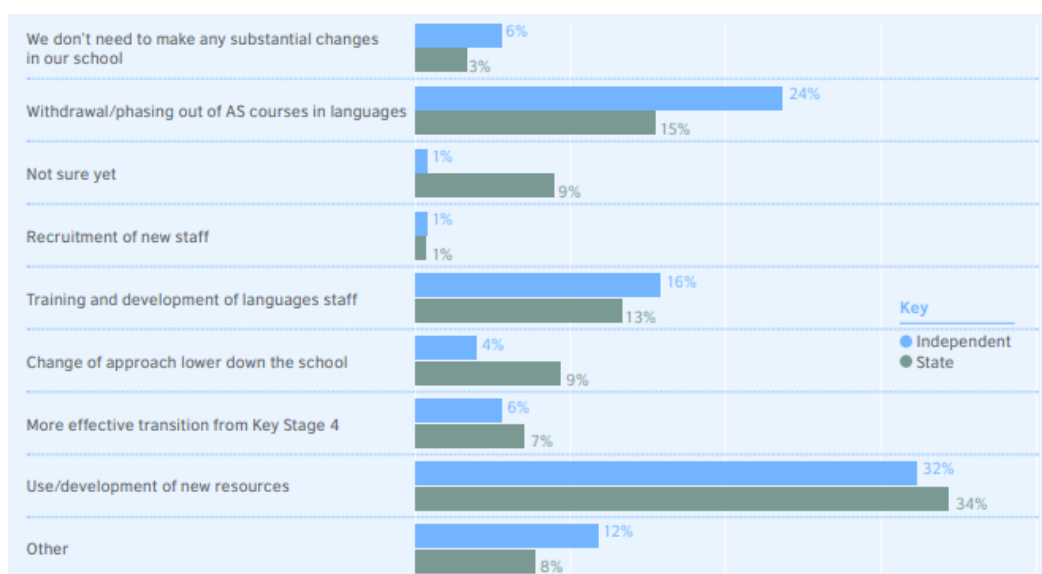


Figure 13 Changes being made in response to the new A level specifications for languages, state and independent schools, 2016 (British Council, 2017).

2.3.4 University uptake

One characteristic of the representation made by stakeholders with regard to inter-subject comparability in modern foreign languages is the degree of agreement and co-ordination between representatives of teaching subject communities, and those from higher education. This vocal involvement from some university language departments differs significantly from the situation in sciences, where higher education representatives have generally remained more detached from the debate and the main conduit for concerns about grading severity have been the learned societies.

There is a consensus amongst teacher subject communities and some university subject representatives that severe grading of A level French, German and Spanish is leading to a decline in uptake at key stage 5 and a commensurate decline in university entries, as most institutions require that applicants have studied at least one language to A level to be able to progress to study it at undergraduate level. Other students may wish to study the subject and opt to do it at A level, but be deterred either by their teacher or by indicators of likely performance (such as GCSE grades and mock exam results) from applying to study it at university, choosing instead to focus on a subject in which they are more likely to gain a higher grade (where this is a criteria of admission).

Both groups are in agreement that this decline in university uptake is likely to be extremely damaging for the national interest of the country (particularly in the context of Brexit), but also that it will create further issues within the subject by leading to a shortage of qualified language teachers. Higher education representatives also point to the closure of university departments as a result of declining uptake, and in particular the decreased interest in German, which they attribute in part to the impact of severe grading dissuading students from studying the subject.

The Universities and Colleges Admission Service (UCAS) publish annual reports, which provide data on trends in applications and acceptances within certain subject areas, reported under JAC3 – the Joint Academic Coding System. The subjects within the scope of this report sit within Group R (European Languages, Literature and related subjects). Whilst this group does not exclusively contain French, German and Spanish, as these are the largest entry European Languages and students are generally expected to possess at least one language A level to apply to study modern foreign languages at university, these are a reasonable measure of progression. Figure 14 shows UCAS entries by subject group in 2017.

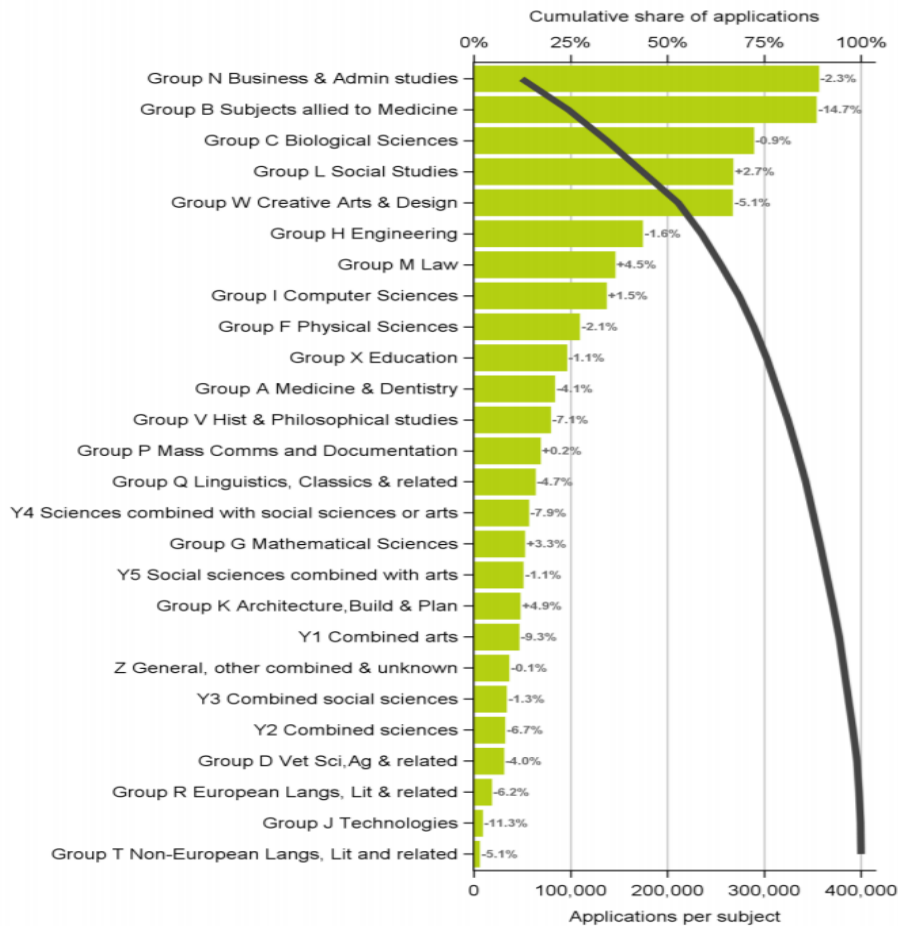


Figure 14

Applications by subject group in 2017 (with cumulative share of applications, and percentage change in applications from 2016) (UCAS, 2017).

[Note bars show number of applications per subject, line shows the cumulative share of applications, and annotations show percentage change in applications from 2016.]

European Languages, Literature and related subjects featured amongst the subject groups with the fewest applicants in 2017, with only 17,970 applications. Most subject areas experienced a reduction in applications in 2017, and the decrease in languages amounted to 6.2%. This is less significant than that in subject such as technologies, combined arts and medicine. This decline is not necessarily evidence of the impact of severe grading (particularly given the claim that this has been a persistent issue and the decline in language uptake at university is relatively recent), but it is illustrative of issues within the subject to which perceptions of difficulty, possibly resulting from severe grading, may be contributing.

UCAS also report data on the proportional change in acceptances by subject group over a period of ten years, from 2008 to 2017. This data is shown in Figure 15.

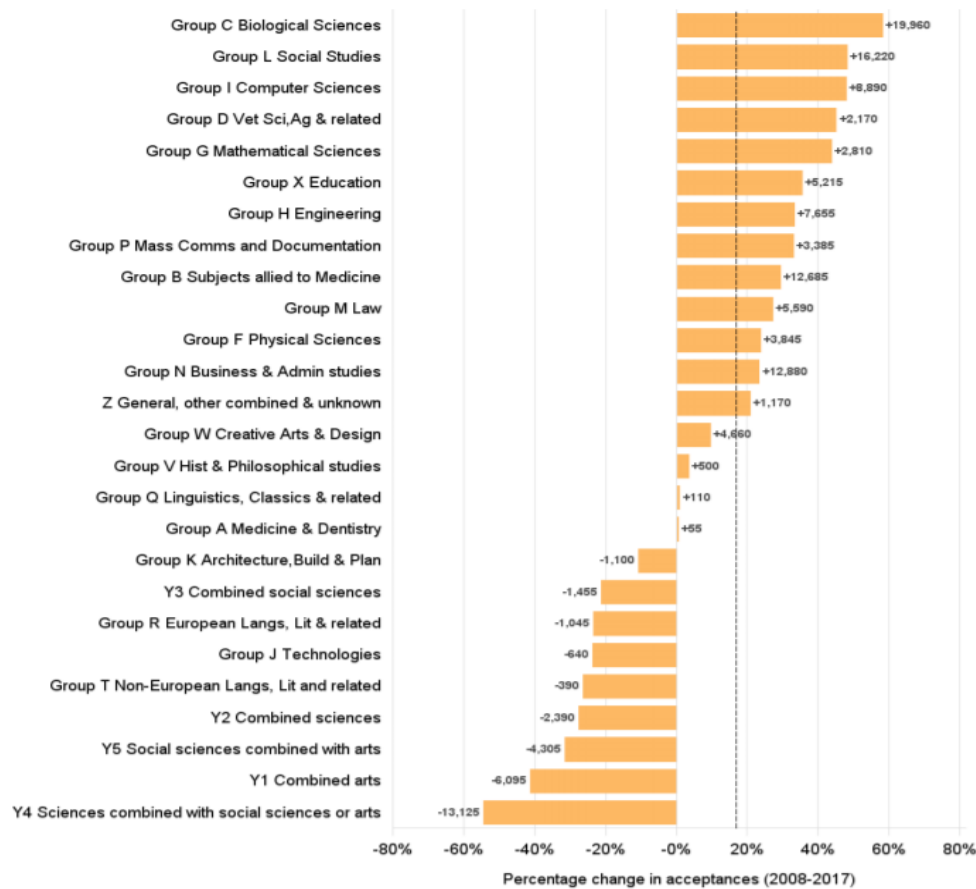


Figure 15 Proportional and numerical change in acceptances by subject group, 2008 to 2017 (UCAS, 2017).

[Note that the dotted line is the proportional increase for all subject groups (+17%). Annotations show the numerical change in acceptances between 2008 and 2017.]

Acceptances to study European Languages, Literature and related subjects are approximately 24% lower than in 2008, with a decline of 1,045. Given that the decline in applications in 2017 alone was 1114 students, this does not suggest that decrease in A level uptake has resulted in a significant increase in the number of positions on university degree courses going unfilled – but rather that universities have not been as selective in their recruitment. Whilst the impact on individual universities might be quite significant, a decrease on this scale does not necessarily suggest a crisis in the system, but may potentially be distorted by increases in acceptances in the years 2008 – 2017. This should be considered in the light of the general increase in acceptances across higher education as a whole.

A more significant decline is shown by aggregating the data related to the various single and combined honours languages courses studied at university. Figure 16 shows that the number of undergraduates studying languages has decreased every year since 2011/2012 – falling by 34% overall by 2016/2017.

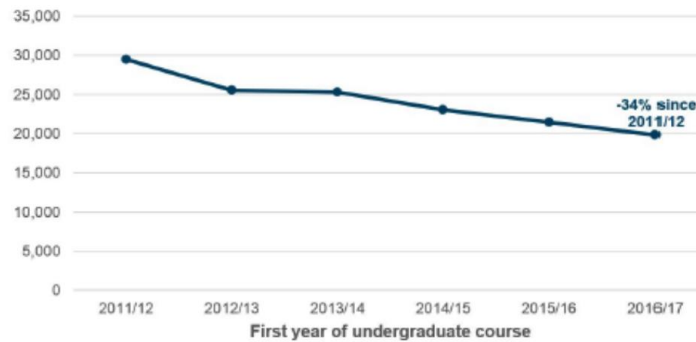


Figure 16 Language applications and acceptances for undergraduate degrees 2011/12 – 2106/2017 (Source: UCAS applications and acceptances for types of higher education course – 2016 (Groups Q, R, T Y)).

On the other hand, decreasing A level entries may not necessarily have led to an overall decrease in the quality of candidates accepted onto university courses if this decrease has been offset by a decrease in the number of undergraduate degrees being offered in universities. Stakeholders indicate that university language departments have received decreased funding or even been forced to close as the number of students progressing from A level to higher education to study modern foreign languages has decreased.

The presentation given by Wyburd (2016) on behalf of the University Council of Modern Languages illustrates the impact on higher education through data which shows the recent decrease in the number of universities offering single and joint honours language degrees. This is shown in Figure 17.

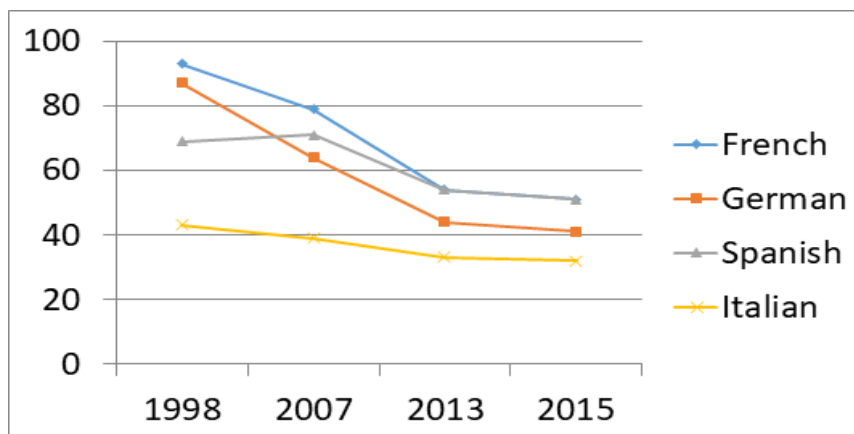


Figure 17 Number of universities offering single/joint honours languages 1998 - 2015 (University Council of Modern Languages, 2016).

The decline is severe, and especially so in German where the number of universities offering single and joint honours degrees in the subject has fallen to less than half the number in 1998. The decline in institutions offering French is also significant, though slightly less. The proportional decline appears to have been greatest in the period 1998 to 2013, and less so in recent years. This provides an interesting comparison with the decline in A level entries for these subjects reported by the JCQ, where the decrease appears to have been reasonably consistent over this

period. This suggests that whilst declining A level uptake may have had a significant impact on the number of universities offering undergraduate French and German, this is now being mitigated to some extent by other factors (for example, if universities were to decide to low entry requirements).

It is worth noting however that this data shows that a significant number of Spanish courses also ceased to be offered over this period – during which Spanish A level entries were in fact increasing. Whilst the decline in the number of institutions offering single and joint honours languages degrees was far less than that in French and German, and the numbers even increased overall 1998 – 2007, the fact that Spanish courses have also been rolled back within higher education suggests that declining A level entries are not the sole (or perhaps even most significant) cause of the waning status of languages at universities.

2.3.5 Evidence of filtering (selection and self-selection)

Tom Allen (2016), data analyst at the Institute of Physics, has sought to illustrate the effect of self- and school-imposed selection due to perceptions of subject difficulty by using 2014 entry data from the National Pupil Database to investigate the number of school sixth forms where only students obtaining GCSE A* and A in a subject progress to study it at A level. His findings are presented in Figure 18.

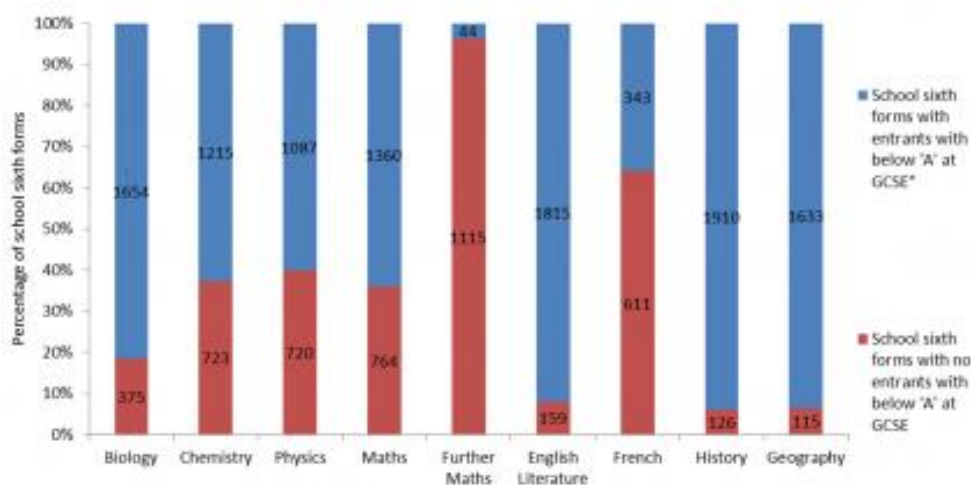


Figure 18 Percentage of school sixth forms with no A level entrants obtaining less than a GCSE grade A (Institute of Physics, 2016).

Whilst intended to illustrate the impact of self- and school-imposed on A level Physics (and to a lesser extent Chemistry and Biology), this analysis showed a more significant number of centres entering students for A level French with no entrants possessing prior attainment below grade A at GCSE– suggesting that the filtering effect in the form of entry criteria may in fact be more significant in A level languages than it is A level sciences.

Further insight into the possible effects of selection upon the cohort of students sitting A level French, German and Spanish can be gained by looking at trends over

time in the relative prior attainment profiles of students in these subjects. Were self-selection or school selection taking place, and if as stakeholders contend, this filtering was exacerbating the grading severity of language subjects and contributing to their perceived difficulty, we might expect the range of prior attainment to narrow over time due to relatively fewer A level entries from low prior attainment bands, as less able students opt out of these subjects (or are opted out by school selection policies). Table 9 shows year-on-year trends in GCSE distribution for science A level cohorts and their mean grade in the same subject at GCSE from 2008 to 2016.

Table 9 Mean and SD of GCSE distributions for A level French, German and Spanish from 2008 to 2016 and their mean grade in the same subject at GCSE.

Subject	Year	Mean GCSE		Mean grade in subject	
		Mean	SD	Mean	SD
French	2008	6.89	0.79	7.48	0.71
	2010	6.91	0.79	7.45	0.71
	2013	6.94	0.81	7.45	0.72
	2015	6.91	0.81	7.50	0.70
	2016	6.92	0.82	7.53	0.67
German	2008	6.79	0.83	7.35	0.78
	2010	6.81	0.83	7.33	0.78
	2013	6.82	0.85	7.34	0.78
	2015	6.82	0.87	7.40	0.75
	2016	6.87	0.83	7.37	0.77
Spanish	2008	6.78	0.83	7.41	0.77
	2010	6.79	0.85	7.44	0.75
	2013	6.76	0.87	7.47	0.72
	2015	6.73	0.87	7.50	0.69
	2016	6.75	0.84	7.48	0.70

In general, students studying MFL at A level are high attaining students, averaging just under a grade A for all GCSEs, and between a grade A and A* for their GCSE language. Mean GCSE has remained broadly stable for French and Spanish, with a slight increase in German. The mean GCSE grade in subject has increased very slightly for Spanish and French, with the SD also contracting slightly for Spanish. This may indicate that Spanish cohort has become slightly more selective.

Stakeholders describe a particular issue affecting the cohort studying A level French, with less able students being dissuaded from studying the subject altogether. Figure 19 shows the grade distribution of students taking A level French 2002 – 2015.

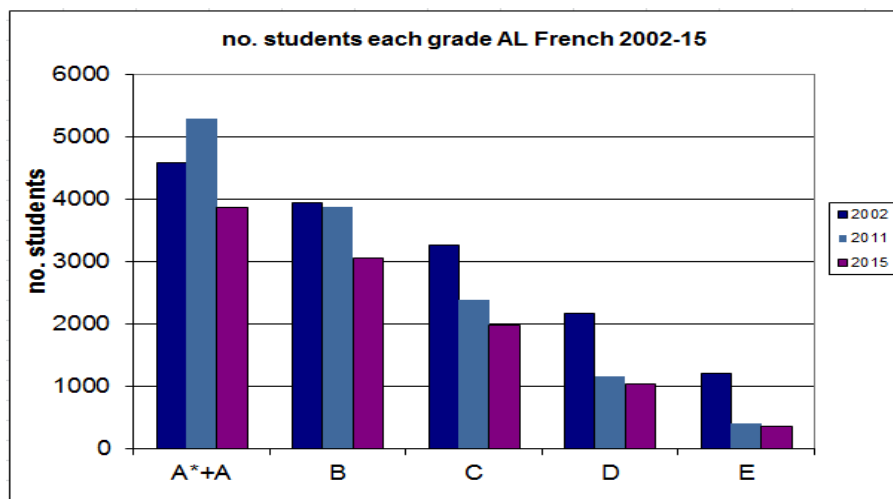


Figure 19 Grade distribution of students studying A level French 2002 - 2015 (ASCL, 2017).

Stakeholders point to the sharp decline in candidates at E grade in particular over the period 2002 – 2015 as evidence that there is now a mismatch between the ‘difficulty’ of the exam and the cohort profile. The number of E grade candidates in 2015 was 29% of the number in 2002. Stakeholders argue that this has led to lower parts of the grade distribution being filled by “downgraded” mid-attaining students, treating them more harshly than students in other subjects and reinforcing the effect of grading severity.

2.3.6 A level entry gender breakdown

Some stakeholders have suggested that perceptions of subject difficulty are likely to discourage certain groups of students from subjects which potentially more severely. Whilst this is primarily a concern of stakeholders in the sciences, we have also considered whether evidence exists that this may be having an impact on uptake in A level French, German and Spanish.

Given that languages also appear to be potentially more severely graded under statistical measures of subject difficulty such as Rasch and CPA, we can also examine the validity of the claim that perceived difficulty is a deterrent to girls in light of the evidence of gender uptake in A level French, German and Spanish to determine whether this argument is plausible.

JCQ have published data on gender participation by A level subject since 2007. Figure 20 shows the gender profiles of the cohort sitting A levels in 2017.

In 2017, A level French, German and Spanish were all more popular with girls than boys. Furthermore, all three also exhibited a gender uptake which was more female dominated than the overall cohort profile.

A level French possessed the highest percentage female intake in of these languages in 2017, with girls constituting approximately 70% of the entry.

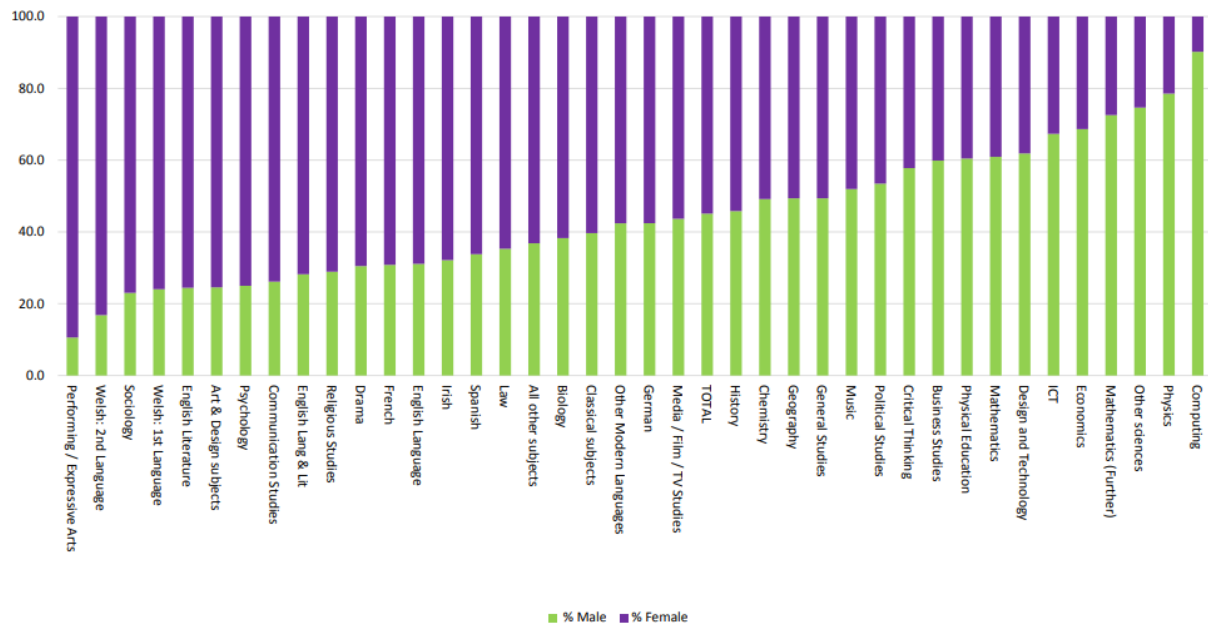


Figure 20 Differences between male and female A level subject choices 2017 (Joint Council of Qualifications, 2017).

Fewer girls took A level Spanish in 2017, but the gender profile of the cohort was still overwhelmingly female. Girls constituted approximately 64% of those choosing to study the subject.

German had the most gender balanced A level entry in 2017, but more girls still chose to study the subject than boys, making up approximately 58% of the entry.

In all three languages, the female-to-male ratio was greater than that seen in the cohort overall. These figures are interesting not only because they indicate that perceived difficulty does not appear to be having a particular deterrent effect on the rates of female entry in languages as is claimed in other subjects overall, but also there does not appear to be any relation between apparent difficulty under statistical measures and the proportion of females studying these subjects – for instance, French appears to be the subject which is ‘hardest’ under Rasch analysis, but has the highest proportion of female students of the three languages.

Finally, within languages this raises additional questions about why entry is so heavily gendered, and what the implications might be for uptake if the subject were as attractive to boys as it is for girls. Much is made of the efforts to increase female participation rates in studying heavily ‘male’ subjects such as physics and computing, for instance through greater exposure to positive role models – within A level languages however the focus of stakeholders is primarily on severe grading as a cause of declining entry, not the consideration of potential deterrents upon certain groups of students.

2.3.7 Teacher supply data

One of the recommendations made in the advice submitted to the Ofqual Board by the Standards Advisory Group³ was that, in addition to evidence from statistical measures of difficulty and evidence of the concerns of stakeholders, the basket of evidence considered to determine whether an adjustment is warranted in a particular subject should also include contextual data such as figures on teacher supply.

Stakeholder concerns about teacher supply within the subjects included in the scope of our work on inter-subject comparability are very different. In A level French, German and Spanish, uptake is declining significantly and has been for some time, with stakeholders expressing fears that this is contributing to a shortage of teachers qualified to deliver the subject (particularly dual linguists), which in turn will lead to fewer schools offering these subjects at A level. They are also concerned that a teacher shortage will contribute to the perceived 'difficulty' of A level French, German and Spanish through a decline in overall teaching quality (though it should be noted that the interaction between teacher shortages and department closures will only lead to a decline in teaching quality if less qualified teachers are appointed to roles that more qualified teachers have vacated: if fewer schools offer the subject but the remaining schools have similarly qualified teachers, then this will have no overall impact on teacher quality). These concerns are more immediate than those seen in the sciences, due to the rate at which entries are declining. In this case, we might consider declining numbers of teachers to be possible evidence of a negative impact on A level French, German and Spanish resulting from severe grading.

Definitive figures on the overall number of teachers in the education system within a specific subject area are unavailable. Furthermore, analysis of teacher supply would also require us to consider factors such as retention rates within a particular subject to determine whether there is an adequate number of qualified teachers. Nor is it clear what other factors would need to be taken into account if seeking to establish whether there had been an increase or a decline in teacher quality – would we also need to take into account education level, main degree subject, teaching experience and pedagogic ability? Is a PhD holder likely to be a more effective teacher than someone with only an MA? Is someone who has been teaching for twenty years likely to be a more effective teacher than a recent graduate?

However, the Department for Education does publish data relating to Initial Teacher Training recruitment which might give us an indication of the number of new teachers, if not their quality. The DfE uses the Teacher Supply Model (TSM) to estimate the number of postgraduate trainees required in England in each subject in a given academic year. This estimates how many Newly Qualified Teachers are

³ The Standards Advisory Group is a committee of the Ofqual Board that reviews research and makes recommendations about how we can maintain standards of qualifications. It is made up of members of our board and independent assessment specialists.

needed to maintain the stock of qualified teachers, taking into account projections of the pupil population, the effect of new policies, and estimates of teacher retention. The number of NQTs needed is then scaled up to account for trainees who do not complete their training and trainees who do not go on to secure teaching employment. This figure is used to calculate the total number of postgraduate ITT places needed. As such, data about recruitment against TSM targets should give us some evidence of the supply of teachers to deliver a specific subject.

Figure 21 shows new entrants to ITT courses as a proportion of TSM targets in 2015/2016 by subject. Whilst teacher recruitment targets under the Teacher Supply Model were not met, the shortfall of only 5% compares favourably with subjects such as physics and also with mathematics – a subject which appears to be more severely graded than French, German and Spanish under statistical measures of difficulty such as Rasch, and which is experiencing issues with teacher recruitment, but which does not attract the same kind of concerns from stakeholders nor show a declining entry.

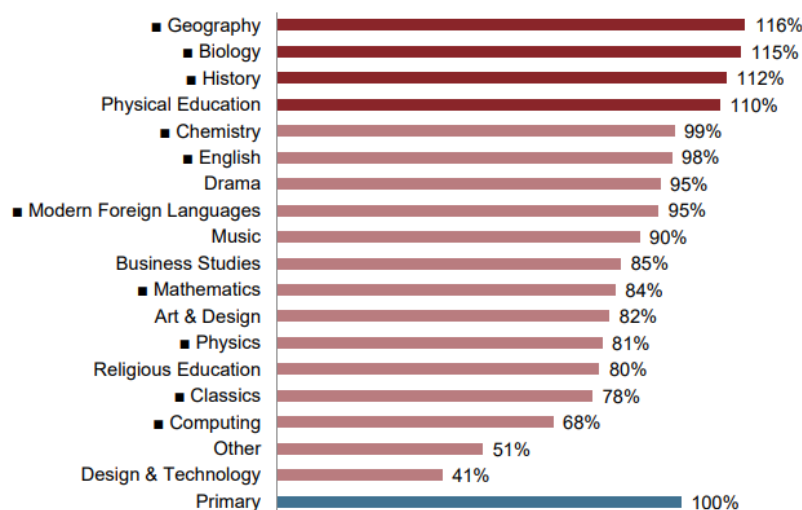


Figure 21 New entrants to

postgraduate initial teacher training by subject, as a proportion of TSM targets, 2015/2016 (Department for Education, 2017).

National College for Teaching and Leadership also published data on recruitment against Teacher Supply Model targets year-on-year. Figure 22 shows the number of modern foreign language postgraduates recruited to Initial Teacher Training courses compared to targets in recent years, with the targets in light blue and the actual recruitment figures in dark blue.

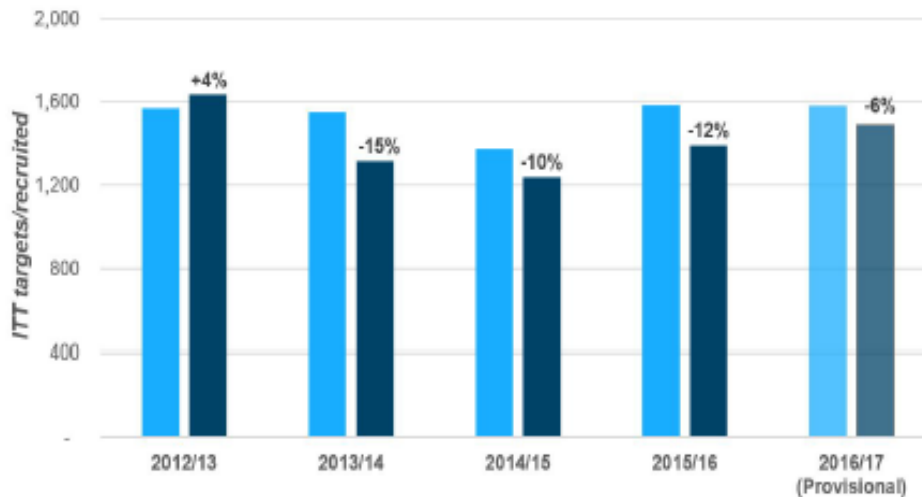


Figure 22 Postgraduate modern foreign languages Initial Teacher Training new entrants against Teacher Supply Model targets 2012/2013 – 2016/2017 (National College for Teaching and Leadership, 2017).

The data shows that teacher recruitment targets for modern foreign languages have not been met for the past five years. Whilst this may not be immediately relevant to the consideration of potential grading severity, it is significant in that the main policy levers possessed by the DfE for increasing teacher recruitment have already been deployed: ITT allocations for modern foreign languages are currently uncapped, and teacher trainees in these subjects already receive the near-maximum bursary of £25,002, regardless of degree class. This may serve to eliminate some of the other potential explanations of the decline in the number of languages teachers beyond the impact of severe grading on the number of dual linguists.

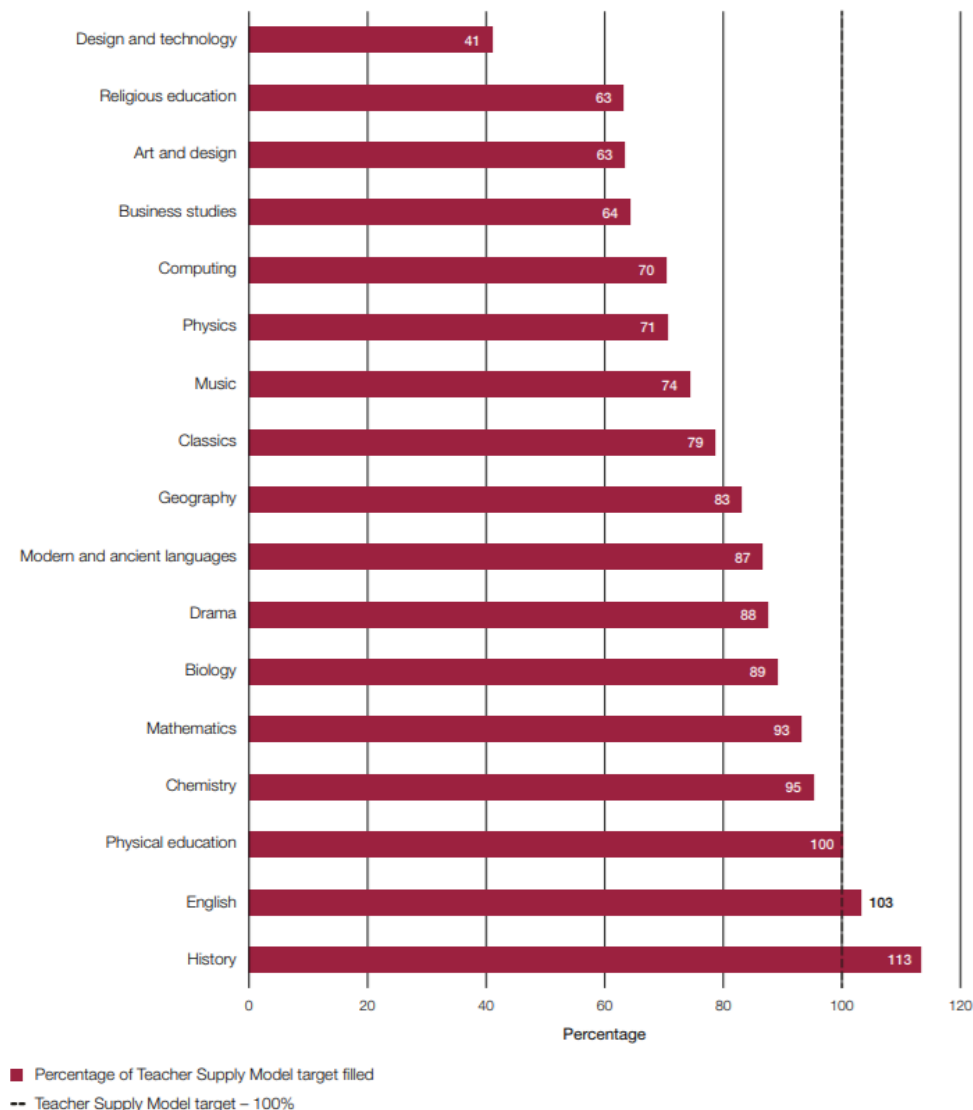
Stakeholders have also pointed to alternative sources of data on teacher recruitment, and in particular to the consideration given to Teacher Supply Models targets by the National Audit Office (NAO). Published in 2016, the NAO report *Training new teachers* sought to establish whether the Department for Education is achieving value for money through its arrangements to train new teachers.

The report identifies some shortcomings in the Department's teacher supply model, stating that whilst it "has strengths... [it] may still inaccurately predict schools' need for trainee teachers", and that "the Department is yet to demonstrate how accurate the model and its own judgements are" (Morse, 2016). The NAO report concludes that indicators suggest teacher shortages are growing, and that the number of classes in secondary schools who are taught by teachers without a relevant post-A-level qualification in that subject is growing. The report specifically identifies Spanish and German as subjects likely to be delivered by a non-specialist, with 43% and 25% respectively of lessons being delivered by teachers without relevant post-A-level qualifications.

The NAO report provides data on the total number of training places filled against targets by subject which differs from that reported from the Department. The NAO uses as its source figures from the National College for Teaching and Leadership and a different estimation of the rate at which the school age population will increase and thus the demand for teachers. Figure 23 shows the number of training places filled against targets by subject in 2015/2016 according to the NAO.

Figure 23 Training places filled against targets by subject, 2015/2016 (National Audit Office, 2016).

There is wide variation in the proportion of places filled by subject



Source: National College for Teaching and Leadership

These figures suggest that the situation for recruitment of languages teachers is actually worse than that reported by the Department, with only 87% of the Teacher Supply Model targets being filled in these subjects rather than 95% per the DfE figures. It should be noted that the NAO data reports modern and ancient languages together, so shortages in other subjects may be impacting detrimentally on the figures for A level French, German and Spanish in this instance. Nevertheless, entry in other A levels languages is very low, so this is unlikely.

As outlined above, some stakeholders have claimed that the decline in uptake for languages at A level, which they attribute to severe grading, is contributing to difficulties with teacher recruitment in these subjects and will exacerbate the issue if prompt action is not taken to adjust grading standards.

The NAO have also attempted to measure the quality of teachers being recruited under the Teacher Supply Model, using the percentage of post-graduate entrants with a 2:1 degree or above as an indicator of subject knowledge. Although this

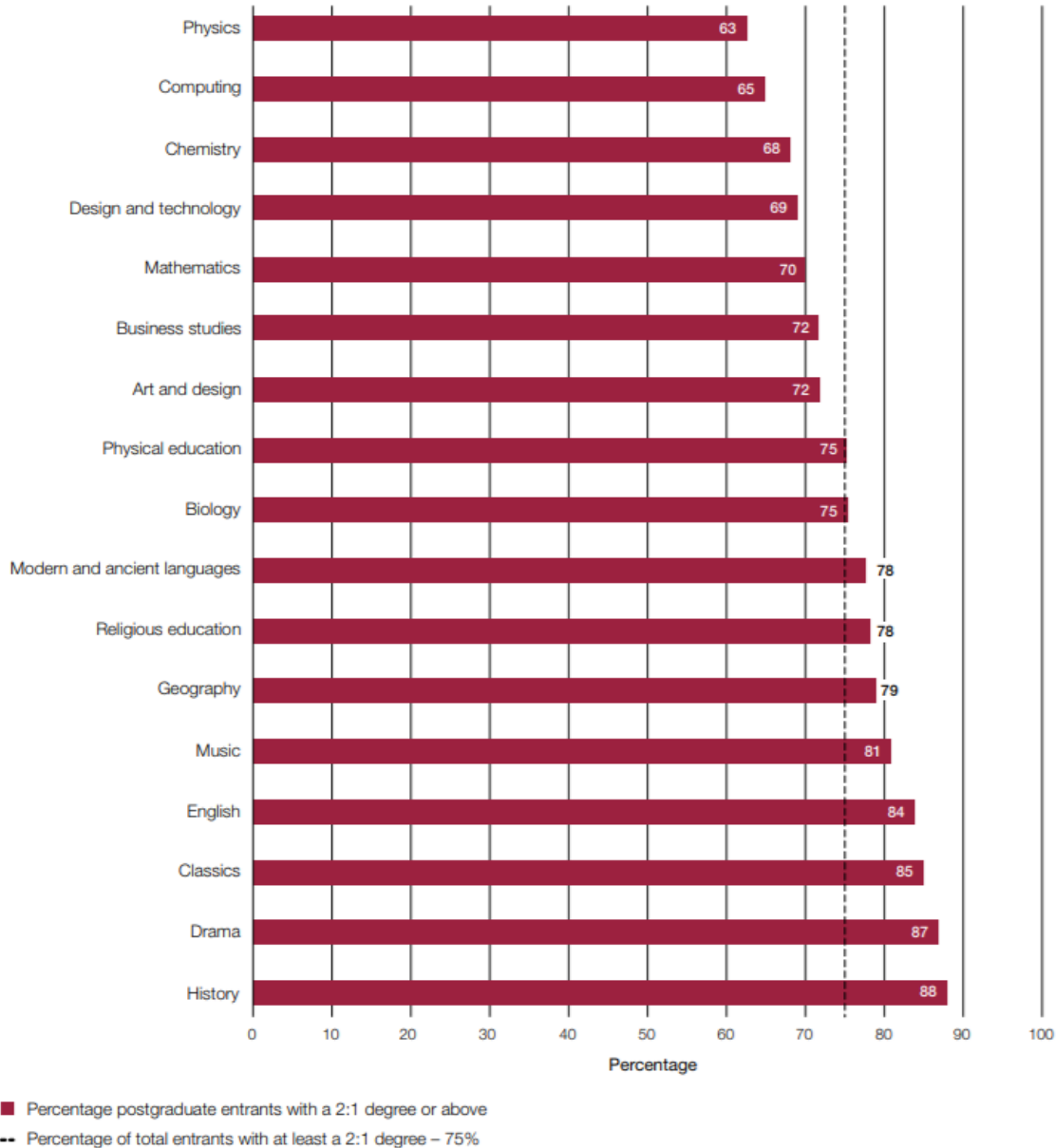
should not be assumed to be a measure of likely teaching ability - recognised in the report, which explains that “schools and providers told us that degree class was a reasonable indicator of subject knowledge but a weak predictor of other aspects of teacher quality” (Morse, 2016) and points to the Sutton Trust publication *What makes great teaching?* which concluded that research on the link between degree class and student learning was “generally inconsistent and hard to interpret” (Sutton Trust, 2014) - it does facilitate comparison between different subjects in terms of the supply of suitably qualified prospective teachers at the beginning of their training, and allow us to consider whether any relation exists between claims of severe grading and another potential aspect of teacher supply. Figure 24 shows the proportion of postgraduate teacher training entrants with a degree class of 2:1 or above in each subject in 2015/2016.

For languages (again with the caveat that the National Audit Office figures jointly report modern and ancient languages, rather than just modern foreign languages), 78% of postgraduate entrants to teacher training possessed at least a 2:1 degree or above. This is greater than the overall proportion of postgraduate trainee entrants with a least a 2:1 degree, which for all subjects was 75% in 2015/2016. This overall figure increased from 63% in 2010/2011, at a rate which exceeds the change in the overall proportion of all graduates receiving upper-second and first class degrees.

Not only is this figure of 78% above the overall average, but it also compares favourably with the proportion of postgraduate teacher training entrants possessing 2:1 degrees or above in subjects such as physics (63%), computing (65%) and chemistry which appear to be ‘harder’ than languages under Rasch and CPA analysis, but also with those which sit close to A level French and German in terms of difficulty such as mathematics (70%), and even those which appear to be significantly ‘easier’ in statistical terms than languages such as art and design (72%) and business studies (72%). This suggests both that other subjects are suffering more severe teacher shortages without it ultimately impacting on uptake (as with chemistry, for which entries are steadily increasing), and also that the causal link between potential severe grading and issues with teacher recruitment is not proven by this data.

Figure 24 Entrants with at least an upper-second degree by subject, 2015/2016
(National Audit Office, 2016)

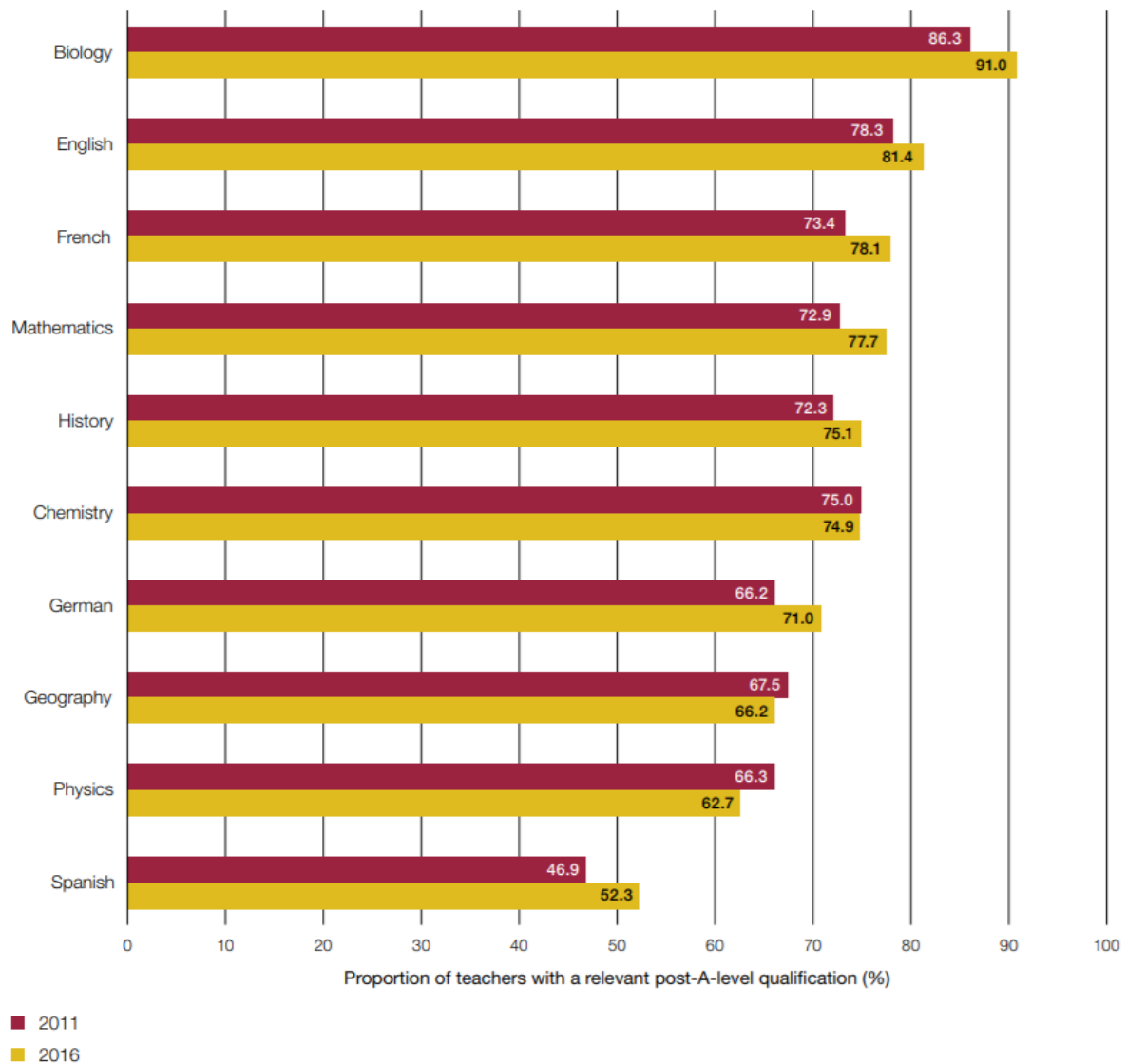
There is wide variation between subjects in the proportion of entrants with at least an upper-second degree



Source: National College for Teaching and Leadership

In 2017 the National Audit Office revisited the issue of teacher supply shortages in the report *Retaining and developing the teacher workforce*. This paper considered the most commonly taught languages as individual subjects rather than as a group, and allows us to examine in greater detail the claims by stakeholders of a potential negative impact of severe grading on teacher quality (rather than just teacher recruitment) in relation to A level French, German and Spanish. The report provides comparative figures which demonstrate how the proportion of secondary school teachers possessing a post-A-level qualification changed between 2011 and 2016, shown in Figure 25.

Figure 25 Proportion of secondary school teachers with a post-A-level qualification and teaching in the relevant subject, 2011 and 2016 (National Audit Office, 2017).



The NAO data shows that within all three languages considered within the scope of this report the number of teachers with a relevant post-A-level qualification increased rather than decreased in the period from 2011 to 2015: rising from 73.4% to 78.1% in French, from 66.2% to 71.0% in German, and from 46.9% to 52.3% in Spanish. It could be argued that this is an expected result of the decline in entry and the subsequent decrease in the number of schools offering languages – hence leading to more of a “buyer’s market” for languages teachers, and an increase in the proportion of qualified teachers within those schools that still teach these subjects. However, this is contradicted by the trends observed in A level Spanish, which is experiencing both increased entry and an increase in the number of teachers with a relevant post-A-level qualification. Certainly, the evidence of the National Audit Office does not support the claims of stakeholders that declining entry for A level languages due to persistent grading severity is having a detrimental impact on teacher quality, although it may be having an impact on teacher supply.

2.3.8 Modern foreign languages pedagogy review

In November 2016, the Teaching Schools Council published the *Modern Foreign Languages Pedagogy Review*. The review, chaired by Ian Bauckham, presents the findings of an investigation into language teaching in secondary schools at a time when many schools are “restricting their languages curriculum... more often than not driven by small or falling pupil numbers” (Bauckham, 2016), and notes the “marked historical decline” of uptake in A level French and German which is mirrored by a similar decline in language study at university.

This review took into account written evidence from 14 stakeholder organisations, and interviews with 33 Heads of Department and 58 teachers from across 33 different schools, as well observations from school visits, an extensive literature review, and international evidence supplied by 14 countries. The Advisory Group for the review included representatives from higher education, language experts, past presidents of ASCL and ALL, and advisors to the government.

One of the recommendations of the review was for there to be “a review of grading approaches, particularly at A level” in light of “evidence from many teachers and schools to suggest that the chances of a pupil with... a B grade in modern languages GCSE achieving the same grade at A level are much lower than a pupil with a B grade in history, English or geography”. Directed at Ofqual and the exam boards, this acknowledged the fact that “the apparent difficulty of achieving a high grade at A level in languages could lead to schools advising pupils against taking a language A level... potentially further exacerbating the shortage of highly qualified linguists.”

However, it should be noted that a further fourteen recommendations in the review relate to quality of teaching, ranging from more effective use of the target language in the classroom, to better selection of textbooks and the need for subject specific curricula for initial teacher training in languages, which are intended to tackle the “decline in take-up of languages beyond KS3 [key stage 3] by ensuring that all pupils are engaged and challenged by consistently good teaching”. Clearly then, there is a need to consider the extent to which some elements of teaching practice may also possibly be impacting on pupil performance in the subject, increasing the perception of difficulty whilst also contributing to the apparent ‘severe grading’ seen under statistical measures such as Rasch and Comparative Progression Analysis.

The review also references Ofsted’s report *Key Stage 3: the wasted years?* (Ofsted, 2015) in which language teaching was found to be unsatisfactory in just under half of the lessons observed, and which was identified in part as contributing to students’ decisions not to study the subject at GCSE and beyond. This was in spite of a recognition on the behalf of pupils of the value of language learning, and was attributed to both a lack of enjoyment of the subject, and the assumption that students required a natural affinity for the subject to be successful.

In addition, the review acknowledges a number of other challenges to effective language teaching including a trend in schools to decrease teaching time at key stage 3 to prepare students for GCSEs in subjects which carry greater weight in accountability measures such as English and maths; negative attitudes from parents based upon their own experiences of language learning at school; the increasing ubiquity of English media and the use of English as a *lingua franca*; and the fact languages consistently appear in surveys as pupils’ least favourite subject.

The significance of this review in relation to our decisions over whether to adjust grading standards in French, German and Spanish will be determined by the extent to which we are persuaded that the apparent grading severity under statistical measures of subject difficulty and detrimental trends in uptake are the result of severe grading – or could be possibly be due to or exacerbated by some teaching approaches, or one or more of the other factors identified in the review. This may temper any adjustments we might decide to make, or else lead us to determine that a change to grading standards would not be appropriate as we cannot be certain it would address the cause of the declining entry or the perception of the subject as 'difficult'.

2.3.9 Research into subject choice motivations

As outlined above, one of the concerns of stakeholders in languages is that the perception of difficulty is leading to the filtering out by selection, or implicit or explicit pressure from teachers, of students who may wish to or would benefit from studying French, German or Spanish but are deterred by the belief that they will secure a better outcome in another subject which is graded more leniently. The plausibility of this argument depends on the extent to which we accept that perceived difficulty (and hence outcome) is a greater motivator for students making A level choices than others such as personal interest or career aspiration.

There are several recent relevant sources of evidence which we might consider when attempting to understand the factors which influence subject choice. The first is the report by Panayiotou, Boulden, Newton and Andersson (2017), which asked 1,595 students about the motivation behind their choice of GCSE and A level subjects. The results of the survey are shown in Figure 26.

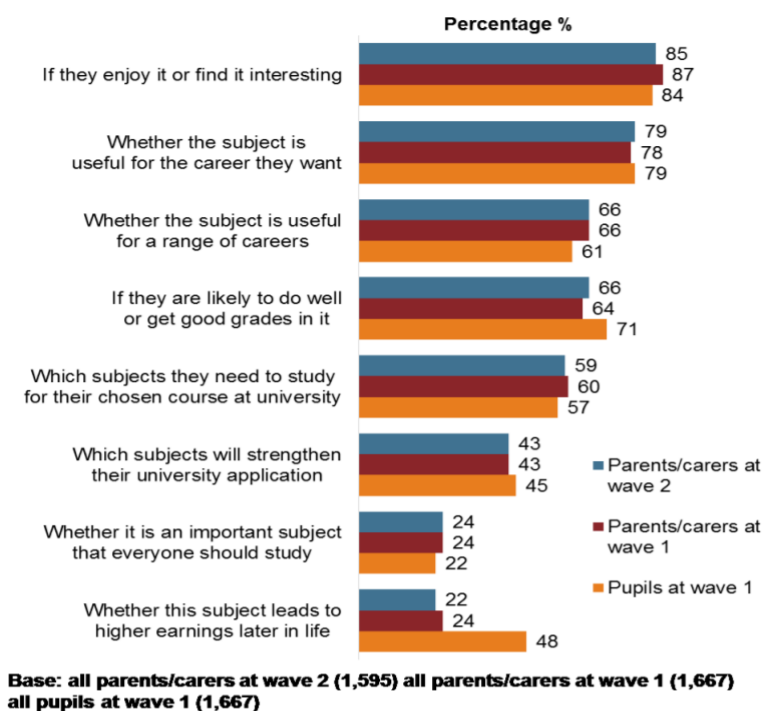


Figure 26 Reasons for choosing GCSE and A level subjects, by survey wave (Department for Education, 2017).

The survey indicates that likely outcome (and by inference, perceived difficulty) is only a tertiary factor influencing subject choice at GCSE and A level. Interest in the subject and future career prospects were both felt to be more significant considerations by students overall. This supports the conclusion of our own research into perceptions of difficulty and the impact on subject choice (see below).

What the survey does not tell us, however, is the number of those of students who did not rate personal enjoyment/interest or career utility as subject choice motivations (15% and 21% respectively) who *did* indicate that they might base their decision on how likely they would be to do well in the subject. Potentially then, even though perceived difficulty may not be the main motivating factor behind the choices made by the sample as a whole, there could be a minority for whom it would have a deterrent effect when considering whether to study A level languages.

Furthermore, these results might look different if GCSE and A level were reported separately. The survey also assumes the decision lies with students, not teachers or schools, who may be deterring students from studying certain subjects by enacting barriers to entry in the form of minimum grade requirements or cease to offer it at all.

Another source of evidence to consider is the research conducted by Cuff (2017). The purpose of this research was to explore whether teachers' and students' perceptions of subject difficulty might be having an effect on which subjects students choose to study in upper secondary education, and whether other concerns (e.g. subject enjoyment or usefulness) might interact with this relationship.

This qualitative survey of 112 students found that perceptions of difficulty were not the main basis of their decisions, and instead their choices were focussed more upon enjoyment and usefulness. Importantly, the majority of students stated that they were often willing to overlook subject difficulty when they enjoyed a subject and/or needed it to satisfy their university or career ambitions – though this leaves a minority who would still be deterred from studying a subject they perceived to be difficult. Students also agreed that although some subjects 'stood out' as seeming to be generally more difficult than others, whether or not they found a particular subject difficult was dependent upon their own individual strengths rather than on any commonly held notions of subject difficulty. Nonetheless students acknowledged that they did base their subject choices on perceptions of difficulty to some extent, but their conception of what made a subject difficult was based on a range of factors including style of assessment, different types of content (including how easy students found them to relate to and understand), workload, and teaching. They also recognised that they were sometimes discouraged by their teachers, parents, and friends from choosing subjects that were thought to be too difficult for them. However, although many felt that they were "discouraged" by their teachers for certain choices, students did maintain that ultimately it was still their choice, and rarely did they feel that they were forced into studying certain subjects.

The survey also found that some students were being prohibited from taking certain subjects at A level by school entry policies which often implemented prevented lower ability (according to their GCSE grades) students from accessing certain 'difficult' courses. These policies were sometimes driven by accountability scores, to ensure that students were able to achieve the highest possible outcomes in the subjects

they chose. Teachers' decisions over which subjects to offer to their students were also sometimes affected by their perceptions of subject difficulty. For example, some schools had elected not to offer certain 'difficult' subjects, again meaning that students were prevented from choosing to study those fields of potential interest. This lends credence to the arguments offered by some stakeholders that filtering is occurring in the entry for A level language subjects as a response to grading severity.

Figure 27 shows the range of influences on students' subject choices, and their interaction, identified in the course of the study.

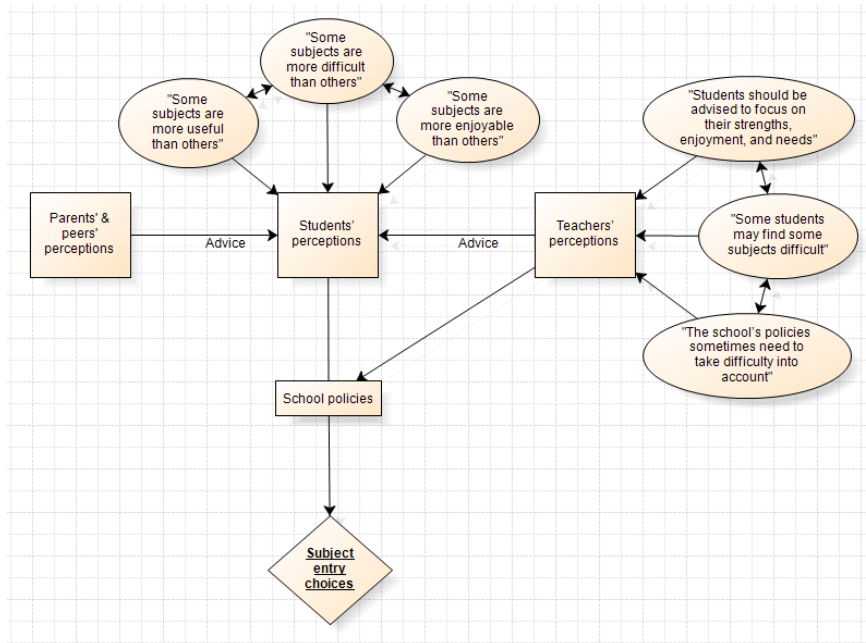


Figure 27 Thematic map of the influences of perceptions of subject difficulty on students' subject choices (Cuff, 2017).

The main conclusion drawn from this research was that subject choices appear to be primarily driven by a triad of perceptions: enjoyment, usefulness, and difficulty (with perceptions being mostly person-specific). Although perceptions of difficulty did have an influence on subject choices, they are perhaps the lesser of these three concerns – with both teachers and students agreeing that choices are more driven by enjoyment and usefulness than perceptions of difficulty (cf Springate et al., 2008, who deemed enjoyment and future ambitions to be 'high influence factors', and difficulty to be a 'low influence factor'). Nonetheless, although this determinant of subject choices is perhaps not as great as previous concerns have suggested, the findings do support the argument that there are some students whose decisions are affected by perceptions of subject difficulty – although these are highly subjective perceptions, which are unlikely to be addressed by an adjustment to grading standards.

Finally, Taylor (2015) conducted research which sought to explore the drivers behind students' subject choice at A level through the psychological theory of planned

behaviour, and which focused specifically on A level physics and media studies⁴. These subjects were chosen in light of concerns about fluctuating entry in physics at a time when A level entries overall, including in those subjects which were perceived by some to be 'less academic' such as media studies, were continuing to grow – leading to fears that students were opting for subjects that were perceived as easier.

The aim of the study was to explore the usefulness of planned behaviour in predicting students' intentions to study AS physics and media studies, and to identify which variables (perceived behavioural control, subjective norms and attitude) were the most important predictors of intentions – as well as the importance of underlying beliefs (for instance, in likely outcome of a behaviour, or of the expectations of others) to these variables. The study was conducted on a random (although not intended to be representative) sample of secondary schools from across the UK who had made entries for AQA GCSE Combined Science, with questionnaires assessing intention to study either AS physics or AS media studies distributed equally amongst final year higher tier GCSE students. Higher tier students were targeted specifically to ensure that the sample included intenders and non-intenders for both subjects. This research was carried out in April to coincide with the time during which students selected their A level options. The questionnaires were completed by 555 students, with a reasonable split by both sex and subject intention.

Overall, the study found that students felt in control of their choice to study physics or media studies, and that they perceived low levels of social pressure from parents and teachers in making their choice. Of the behavioural beliefs tested in the questionnaire, the items that were most strongly related to intention to study were those concerned with the potential positive outcome from taking the subject. Those with higher intention believed that studying physics or media studies would offer more and better career opportunities, and secure them a place on their chosen degree course. Whereas the strength of the correlation between intention and utility for further study or range of job opportunities was the same for both physics and media studies, there was a significant difference between the two in terms of quality of career prospects. Potential physics students regarded studying the subject as more likely to help them secure a good job in the future than those intending to take media studies.

Of the behavioural beliefs concerning less positive associations of study, there was a positive relationship between intention to study physics and the belief that doing so would entail high workload or the study of difficult content. This correlation was negligible for media studies. Similarly, students with higher intentions were also those who perceived greater social pressure and were more motivated to comply with the expectations of others – and again the relationship between normative belief

⁴ *Using the theory of planned behaviour to understand students' subject choices in post-compulsory education* (Rachel Charlotte Taylor, CERP, 2015)

and intention tended to be stronger for physics than media studies, with the difference in the strength of the correlations being statistically significant for those items relating to expectations of parents and teachers. This indicates that students intending to study physics were more affected by external pressures than those intending to study media studies – but also the correlation suggests that this pressure is steering them towards the subject rather than away from it.

Of the control beliefs tested in the study, in both media studies and physics students with higher intentions to study the subject felt more personal control over their subject choice. Again, this is suggestive that those who do wish to study physics are not a particular risk of being influenced away from it by external factors. Comparison of the strength of the relationship between control beliefs and intentions by subject showed no statistically significant difference in the correlation for belief in either how difficult the subject would be or the likely quality of the teaching they would receive – only for how interesting the subject content would be, with the sample again showing a stronger correlation for physics than media studies.

The study concluded that ‘students’ subject choices at A level are most likely to be influenced by their attitude towards studying the subject and the perceptions of significant others, particularly parents and teachers.’ Whilst the report noted that the opinions of parents and teachers were both important in helping form students’ decisions, it was the expectations of parents which were found to be most strongly related to intention to study both physics and media studies – not those of teachers.

The findings of the report that the beliefs most strongly related to intention were those which associated positive outcomes with selecting a particular subject (especially in terms of future career prospects), more so than beliefs about difficulty, support the conclusions of the Ofqual research into the potential impact of relative subject difficulty and the DfE survey of the motivations behind subject choice.

2.4 Views of stakeholders

2.4.1 Subject associations’ views

In making representation to Ofqual on inter-subject comparability in A level modern foreign languages, the subject associations representing language teachers have acted in concert with organisations representing school leadership in both the state and independent sectors. Indeed, one commentator has argued that the fact that the concerns about severe grading have now spread beyond specialist interest groups to organisations with less vested interest demonstrates how serious the impact of severe grading has become. A joint letter, signed by the Association of School and College Leaders, Association for Language Learning, Independent Schools’ Modern Languages Association and the Headmasters’ and Headmistresses’ conference, was

sent to the Chair of Ofqual, Roger Taylor, in July 2017 (Barton et al., 2017). Additional representation has been made in various meetings between Ofqual staff and individuals from ASCL, ALL and ISMLA; in written correspondence; and in presentations delivered at events including the 2015 Ofqual conference on inter-subject comparability.

These stakeholders have all welcomed the ongoing commitment of Ofqual to investigate potential severe grading in A level French, German and Spanish, and the extent to which they have had the opportunity to contribute to the ongoing debate on the topic. However, they have also stressed that they feel that there is an urgent need for action to be taken as soon as possible, and that the negative trends in A level language uptake have intensified to the point where “languages are in danger of being lost in our curriculum” (Barton et al, 2017) at the very point where broader political developments mean that they are becoming increasingly vital to the UK. Aside from the “educational personal, social and cultural” benefits of language study, stakeholders also stress their economic importance – citing studies from the Confederation of British Industry which show that not only do many businesses “want employers to have better ability in languages... [but] it’s estimated that this lack of ability costs the UK 3.5% of economic performance.” Brexit in particular is identified as a significant factor which increases the demand for capable linguists: “Brexit makes them more, not less important – guaranteed trading relationships with Europe may be harder and as a result we will need to forge and maintain relationships... so Brexit increases the importance of languages.” This, they argue, means that the government and Ofqual should be actively seeking to encourage uptake of the subject – not dis-incentivising it through severe grading, as they feel is the case now.

In addition, these stakeholders point to the longstanding nature of concerns about severe grading in A level German, French and Spanish; citing the work of Peter Downes and the Nuffield Report, and the Dearing Review of 2007, as corroborating their belief that these subjects have been out of alignment for some time. In particular, they refer to the conclusions of the Dearing Review which “found strong confirmation of the view that award of grades [in languages] is more demanding than for most other subjects” (Dearing, 2007). The subject associations have also argued that the scale of the problem is conveyed by the fact that, of the 52 subject-specific responses which were received in response to the consultation on potential policy positions which accompanied the publication of the Ofqual working papers on inter-subject comparability in 2015, 28 of those came from the modern foreign languages teaching community.

With the studies outlined above, and the publication of Ofqual’s own inter-subject comparability working papers investigating the apparent difficulty of different A level subjects through statistical measures of subject difficulty such as Rasch and Comparative Progression Analysis, stakeholders now feel that the case that these subjects are more severely graded than others has now been established. They find this evidence to be compelling and conclusive proof of their suspicions that candidates in A level French, German and Spanish were not being treated the same as those in other subjects. They also hold that the debate should now move on from discussing the ‘perception’ of severe grading, given that statistical evidence demonstrates that this is the case, and on to how best to address the “historical anomaly that means that students are achieving on average half a grade lower on their languages qualifications than in all other qualifications” (Barton et al, 2017).

The subject organisations also believe that there is a false logic in the conclusion reached in Ofqual's previous work on inter-subject comparability, that difficulties in conceptualising comparability (and potential challenges to the assumptions of Rasch and Comparative Progression Analysis of a 'general academic aptitude' as a linking construct which allows comparisons to be drawn between different subjects) means that we should not act to achieve alignment on a statistical basis. Rather, the subject associations feel that in light of the statistical evidence and the detrimental impact severe grading may be having on the subject, that "although the concept of comparability may be elusive, its manifestation in the measures outcomes for pupils should have historical anomalies addressed to ensure that pupils (and schools) are not disadvantaged by their curriculum choice (Barton et al, 2017)." They feel that, in the absence of a systematic approach to inter-subject comparability, that it is as equally valid to state that there is no justification for the status of A level French, German and Spanish as outliers under Rasch analysis and other statistical measures of subject difficulty, and that we should seek to adjust grading standards accordingly to "create a level playing field" for students.

Furthermore, one stakeholder has made the argument that even if we consider the potential interactions between severe grading and other previously identified assessment issues which may be having an impact on candidate performance in the subject (e.g. native speakers; poor assessment function; and an incorrectly carried forward A*/A standard – discussed in more detail later in this section) "are now so complex that finding an appropriate new level is challenging" (i.e. that we are not convinced on the basis of statistical evidence alone to look to realign grading standards), that we should be "prepared to consider the health of the subject in addition to issues solely related to grading" (Mair, 2017). Regardless of the evidence to support or refute the link between alleged grading severity and the decline in uptake, "the situation surrounding MFL is now so severe that significant action is required" – and that the consequences of not intervening, which threaten the continuing viability of the study of languages post-16, outweigh any conceptual objections we might have.

This argument rests on the combination of statistical evidence of severe grading suggested by measures of statistical difficulty such as Rasch and Comparative Progression Analysis (as proof that an issue of alignment exists per se) and the declining uptake of the subject at A level and higher education (amongst other negative impacts) as validation of the hypothesis that statistical evidence suggests. The statistical measures of subject difficulty may not provide a clear and compelling argument for realignment when considered in isolation, and similarly we may query whether the perceived difficulty of A level French, German and Spanish is indeed wholly or even partially responsible for declining entry. Taken together however, the subject associations are of the view that this presents both a plausible explanation of the waning status of languages and evidence of the negative impact severe grading is having upon the health of the subject: *quod erat demonstrandum*.

Stakeholders are also firmly of the view that students are being dis-incentivised from the study of languages. This filtering takes two distinct forms.

Firstly, students are dis-incentivised directly in that they perceive from GCSE (which subject associations also believe is severely graded in comparison to other subjects) that they do not perform as well in languages as in other subjects, or are aware of and are drawing comparisons between predictors of likely performance (ALIS,

MIDYIS) in languages compared with other A levels. These relative comparisons lead to students selecting other subjects to maximise their university prospects or UCAS attainment. For previous versions of A levels which have been modular, stakeholders have suggested that this is particularly the case where students studying languages post-16 compare their AS results for French, German or Spanish to other subjects, and decide based on their relative underperformance due to severe grading not to continue to study them at A level. This is something which stakeholders feel has previously been a particular issue in French. Presumably this is due to French possessing the status of most frequently studied language, and hence is more likely to be a student's 'other' A level on which they are not necessarily basing their career prospects but might enjoy enough to include within their options – or else would benefit from studying as a third A level (the implication is that other, less common A levels such as German and Spanish are more likely to be studied by those hoping to progress to study languages at A level, and therefore students are either more likely to stick with them or universities are more likely to make lower offers to fill places).

A level qualifications have now moved from a modular to linear format, something which stakeholders would argue compounds the issue – with students generally taking fewer A levels as schools move away from AS levels (and in doing so are more likely to focus on three 'core' subjects necessary to an identified career path, and less likely to take a fourth subject for interest or enjoyment which might have been a language), students are also conducting the comparisons of likely outcome earlier and ruling out languages immediately after GCSE on the basis of perceived difficulty. With greater exposure to a language beyond key stage 4, these students might have been motivated to study it further at university. Even if they would not have chosen to do so, their loss means fewer students choosing to take the subject and pushes subjects further to the point where they potentially become non-viable: removing the opportunity for others to study the language who may wish to do so.

Secondly, stakeholders make the case that students are being indirectly disincentivised, or even wholly prevented, from studying languages in schools where apparent underperformance due to severe grading (or as a result of declining entry) has led to languages ceasing to be offered at A level altogether; decreased numbers of classes or fewer languages being available as a result of resources being prioritised in more 'successful' subjects; or barriers to entry put in place such as higher entry requirements for progression to study French, German or Spanish at A level.

Interestingly there is little suggestion from language stakeholders that students are being discouraged from taking the subject by teachers concerned about their likely performance in a subject which is apparently severely graded, as there is amongst the science community. This may be indicative of the pressures that A level languages are under in terms of entry that teachers do not feel able to turn away any interested students from the subject regardless of likely outcome, and in turn may be a potential explanation of the apparent severe grading of these subjects under statistical measures of subject difficulty if they find themselves in a position where they are forced to take weaker students to keep courses viable. Given the general trend in declining entry this is unlikely to be a significant factor, but could be present and interacting to some extent with other detrimental factors in A level languages.

There is an additional aspect to consider amongst the possible influences that perceptions of difficulty due to severe grading could be having on the subject. This is highlighted in the British Council *Languages Trend Survey 2016/2017*, and referenced in representation from a languages stakeholder, which reported that “the comparative difficulty of exams in languages in relation to other subjects... are deeply de-motivating for both pupils and teachers.” Decreased motivation could be adding to the apparent difficulty of these subjects (as students resigned to the perception they will be graded more severely and less motivated teachers are likely to perform worse than they would if this were not the case) – and importantly as it is based upon received perceptions rather than necessarily upon direct experience, this could well be occurring irrespective of whether there is any truth to claims of severe grading. Indeed, the *Language Trends Survey* goes on to report that “the exam system is seen as one of the principal barriers to the successful development of language teaching.”

The subject associations all share the conviction that, as a matter of principle, students studying languages should be able to expect to obtain a similar grade in A level French, German and Spanish to what they might receive in other traditional academic (or ‘facilitating’) subjects. Whilst this applies to the qualifications as a whole, they also focus particularly on the number of students achieving A/A* grades in these subjects, which they feel has historically been unfairly depressed (an issue which Ofqual has in recent years taken action to address – discussed in greater detail later in this report). As with the position on severe grading outlined above this is again a long-held view of the subject associations, voiced previously in response to the findings of the Dearing Review. In light of the persistent nature of these concerns, the subject associations argue that the need to address a historical misalignment exceeds concerns about maintaining comparable standards over time: “This is not an issue of dumbing down... it is levelling the playing field with other similar subjects... not to maintain a grading standard that was unfair to begin with (Myers and Blow, 2017).

As well as the evidence of severe grading provided by statistical measures of subject difficulty which these stakeholders feel to be compelling, they also point to a number of negative impacts upon the subject which they also attribute to severe grading – and which, if not addressed, are likely to contribute further to the decline of A level French, German and Spanish; necessitating prompt and significant action. This is a feature of stakeholder views on the inter-subject comparability: that the potential negative consequences of not acting to address perceptions of severe grading in the subject are such that they require an immediate and pragmatic response, rather than a conceptualised one.

Again, some of the factors which stakeholders identify as evidence of the misalignment of standards in A level languages are offered simultaneously as evidence of severe grading but also as issues which will have further negative impacts on these subjects in the future if allowed to continue.

The first of these is the decline in dual linguists observed at A level, and identified in the *Language Trends Survey 2016/2017* (discussed above in section 2.3.2). Whilst the majority of schools reported that the number of students studying more than one language post-16 had experienced no significant change in 2016, 25.6% of state schools and 37.3% of independent schools reported that the figure had fallen – and 14.7% of state schools reported that students no longer had the option of studying two languages due to provision being reduced. Stakeholders feel that this is

particularly likely to have an impact on the number of students progressing to study two modern foreign languages at university, and significantly who might then go on to take the Post Graduate Certificate in Education. Modern Languages teachers are usually expected to be able to deliver more than one language to a given standard of proficiency, and with a decline in the number of students studying two languages at A level the suggestion is either that an increasing number of languages graduates will feel that this career path is closed to them (impacting on teacher supply), or that language teaching in schools will suffer as a result of teachers being required to deliver lessons in a second foreign language which is comparatively weaker than their first. Either of these situations, stakeholders argue, will have a detrimental impact on teaching quality in modern foreign languages as a whole, adding to the perception that this is a 'hard subject' – and may hasten the demise of the subject altogether if departments begin to close because they are deemed either to be unsuccessful or no longer financially viable.

The second factor, which may well be contributing to the first, is the impact of the move from modular to linear A levels as part of the recent reforms. Stakeholders have stated that the overwhelming trend now is for schools to offer three rather than four A level subjects, dropping AS entirely, and this is supported by the findings of both the *Language Trends Survey 2016/2017* and Ofqual's own provisional entry data. Anecdotally, stakeholders identify A levels in modern foreign languages as those most likely to be overlooked by students choosing a core of three A levels, perhaps focusing on likely career progression or on maximising attainment to meet their university offer; whereas previously they might have chosen to study a language as a fourth subject on the basis of enjoyment, interest, or secondary utility for their career (whether by making them more employable or offering the opportunity to live and work abroad). In particular, the subject associations point to the fact that UCAS offers are predicted on the basis of three A level grades as being a key factor. Independent schools, which on the basis of their overall entry might be considered a relative 'haven' for languages, are not exempt from this impact – and those that have already moved to three A levels at key stage 5 are reporting a significant reduction in numbers. For instance, an independent school reported to us last year that the 45 pupils taking AS French in Year 12 would become to 16 pupils taking the A level course in Year 12. This impacts not only on general language proficiency of the country as a whole, something which stakeholders feel should be encouraged, but also progression to university.

A further factor, which stakeholders feel will also contribute to a further decline in teaching quality and hence negatively impact the perceived 'difficulty' of these subject and the health of modern foreign language A level as a whole, is the result of a combination of the shortage of dual linguists and issues with teacher supply. Subject association representatives point to "the under-reported effects of the combination of a shortage of dual linguist teachers and the increase in teachers changing jobs at the end of a term rather than the school year" (Mair, 2017), presumably the result of the increasing competition for teachers as numbers decrease, but also attributed in part due to a less receptive climate for teachers originating from abroad following Brexit. One stakeholder quotes a figure of a "likely shortage of 4,000 modern foreign languages teachers" in 2018, and offers confirmation of the impending shortage of language teacher trainees based on their experience at a prominent institution offering school-centered initial teacher training (SCITT), where they are witnessing first-hand the struggle to recruit. Stakeholders have suggested that the impact of this upon a small subject is significant, and that as "teachers now routinely change job or retire at

the end of a term rather than a year, then existing classes have to be catered for, often by offering a short term post or by employing a Germanist when the real requirement is for a Hispanist. This causes detrimental disruption or the appointment of a teacher who may not be the best suited.” This would be an issue even if teacher recruitment targets such as the DfE’s Teacher Supply Model were being met, as fewer dual linguists would mean that those teachers who were recruited would be less versatile in terms of the languages which they could deliver – and as such is ultimately being attributed to the impact of persistent severe grading, given this would doubly affect those taking two MFL A levels.

Finally, stakeholders also claim that the widely perceived issue of severe grading in A level French, German and Spanish, and in particular the extent to which this influences subject choice, has also had a significant impact on the composition of the A level cohort sitting these subjects. The subject associations argue that there has been a significant decline in the number of weaker candidates entering examinations in A level French over the period 2002 – 2015, with the number of candidates achieving an E grade falling significantly. This they feel has led to a mismatch between the ‘difficulty’ of the examination and cohort profile, which in turn had a knock-on effect on the number of pupils being awarded A*/A grades in 2015 as a result of students being ‘downgraded’ to provide the distribution of lower grades. They are concerned this may also be happening in A level German and Spanish. Though Ofqual undertook work to address this issue through positive awarding tolerances as part of steps to ensure that exam boards would “produce question papers [that] differentiate in a more reliable way between the more able students”, stakeholders still feel that this historic misalignment has not only contributed to the perception of severe grading in the subject, but also continues to pose issues of misalignment with other A levels. In their view Ofqual intervention in these qualifications may have mitigated the historic misalignment to some extent, but the misalignment itself would still remain.

In light of the concerns outlined above, and their conviction that the indications of severe grading under Rasch analysis and other statistical measures of subject difficulty are confirmed by the ongoing negative trends in entry observed within A level languages, the subject associations are convinced that action needs to be taken to adjust grading standards. Whilst the subject communities are clear that they recognise action on severe grading alone will not be sufficient to address the negative trends in A level French, German and Spanish (and evidence relating to the causal links between severe grading and decreased uptake may not be clear), they feel that the potential threat of losing languages entirely from the curriculum of schools and universities is sufficiently compelling to take action on an issue which we can observe and which may be contributing to the problem: “Making changes here [to address severe grading] will not **in itself** solve the overall problem [of declining pupil uptake], but, to use a metaphor, will remove having a hand tied behind one’s back” (Myers and Blow, 2017). An adjustment to grading standards may be only one element of a combined approach to address issues of uptake in the subject (which might also include improvements in pedagogy, scheduled teaching time, and a change in the way the government incentivises language study through performance measures) – but subject associations feel it is a necessary first step.

2.4.2 Perspectives from higher education

As mentioned previously, a characteristic of the stakeholder evidence in relation to inter-subject comparability is the extent to which there is clear agreement between the views of the subject associations and at least one influential section of the higher education languages community. Representation from higher education on the matter of severe grading has been in the form of two letters, one addressed to Ofqual's Director of Strategic Relationships for General Qualifications and jointly signed by two teaching fellows in the modern foreign language faculties of two universities, and the other sent to the Chief Regulator, and co-signed by representatives of the leading modern languages subject associations in higher education in the UK (some of whom also held positions on Ofqual's Modern Foreign Languages Subject Group and the A Level Content Advisory Board which made recommendations on the content of the reformed A level qualifications to the Department for Education). The signatories include the Chair of the University Council of Modern Languages (UCML), the Lead Fellow for Languages at the British Academy, and the Presidents of the Society for French Studies, the Association for German Studies in Great Britain and Ireland, and the Association of Hispanists for Great Britain and Ireland. Meetings have also been held between representatives of these HE subject associations and Ofqual.

In its correspondence, the HE languages community is unequivocal in its support for the subject associations representing teachers and the school leadership associations, and "urge Ofqual to accept that there is now sufficient evidence to demonstrate the severity of grading in MFL... and that the steps necessary to rectify this anomaly need to be implemented with urgency" (Carruthers et al, 2017). These HE representatives welcome the opportunity to contribute to the discussion around potential severe grading in A level French, German and Spanish, as they believe that they are seeing first hand the impact of severe grading on the progression to university study of modern foreign languages as a result of "the decline in Modern Foreign Languages in schools for the past decades... and [are] witnessing the knock-on effects for university departments for university departments of Modern Languages" (Kohl and Mentchen, 2017).

These stakeholders argue that "the evidence of severe grading and its detrimental effects for the health of the subject [were] already considered persuasive by school teachers, university teachers and journalists investigating the issue in 2013" (Carruthers et al, 2017), pointing to the publication of the article *Modern languages: degree courses in freefall* in October of that year in *The Guardian*, which was accompanied by an open letter signed by eighty members of higher education modern language departments and the relevant subject associations (see <https://www.theguardian.com/education/2013/oct/08/modern-foreign-language-degrees-axed>). This they offer as evidence for the widespread support for action in this subject. Despite the recent work undertaken by Ofqual to address assessment issues in A level French, German and Spanish, they believe that there still exists a "fundamental unfairness" in the historic misalignment of standards which means that language A levels are still more severely graded than other subjects, as indicated by Ofqual's own research into statistical measures of difficulty such as Rasch and Comparative Progression Analysis.

Contending that there is now a “broadly based consensus... that choosing a language A level is strategically unwise for any student who wants to secure a good university place” (Kohl and Mentchen, 2017), and that this is leading students to choose other subjects over A level languages (self-selection as a form of filtering, as discussed elsewhere in this report – although they also refer to the need to demonstrate that grading is fair to teacher, head teachers and parents too in recognition of the direct and indirect impact which these individuals can have on influencing a student’s subject choices), the HE subject associations argue that not taking action to address perceived severe grading will “drive even more young linguists out of the subject and make recruitment and retention of MFL teachers even more difficult than it already is”. This is supported by University Council of Modern Language data which shows that, during the period 2002 – 2015, A level entries in French decreased by 34% and in German by 43% (although in Spanish they increased by 56%, this was still a net decrease of 18% when these subjects are combined). The implication here is that those students who might have gone on study languages, and in turn teach them, will no longer choose to enter the system (or will have fewer opportunities to do so if university departments continue to close as a result of declining uptake) and that in turn this will contribute to the shortage of teachers in the subject – increasing the experienced ‘difficulty’ of A levels in French, German and Spanish for those pupils sitting them.

One group of HE stakeholders seeks to make the case that the impact upon A level German has been particularly pronounced, having become “increasingly conscious of the catastrophic impact severe and unreliable grading has had on our subject as a whole, and German in particular” (Kohl and Mentchen, 2017). Data gathered by the University Council of Modern Languages showing the decrease in the number of institutions offering single and dual honours undergraduate degrees since 1998 does seem to suggest that German departments have been particularly badly impacted by the decline in entries – although it is worth bearing in mind that statistical measures of subject difficulty such as Rasch and Comparative Progression analysis suggest that French is more severely graded. Similarly, as pointed out earlier in section 2.3.3 of this document Spanish has also seen a reduction in the number of departments offering the subject, and yet A level entries are increasing overall. Furthermore, the rate of closure of university departments (albeit not the decline in undergraduate entry) appears to have slowed substantially in recent years in all three of A level French, German and Spanish, although it is still a negative trend overall.

Stakeholders suggest that these trends will continue unless confidence in modern foreign languages is rebuilt through action by Ofqual, and that for this to take place there “must be convincing evidence that an able, well-taught and hard-working young person with no prior foreign language competence has had a fair opportunity to achieve the grades A* and A in MFL in the same way as in other comparable subjects” (Kohl and Mentchen, 2017). This argument shifts the burden of proof from establishing that severe grading is occurring in these A levels subjects in order to justify making an adjustment to grading standards, to providing evidence that standards in modern foreign languages are fairly aligned with those in other A levels – a level of certainty which our current conclusions about the usefulness of the evidence of statistical measures of subject difficulty do not necessarily provide.

It is characteristic of the views of stakeholders that they focus on the number of candidates obtaining grades A and A* in A levels in French, German and Spanish. Partially this could be due to the belief that students may be focusing on the subjects

in which they are mostly likely to obtain the best results for university entry generally, or the fact that students are more likely to progress to study those subjects in which they perform best. It may also align with the suggestion from subject stakeholders that schools are likely to regard language departments as performing less well than others if they do not return the same proportion of higher results. If A level French, German and Spanish are indeed more severely graded, this may lead to strategic decisions on the behalf of schools and colleges to reduce their resources or even cease to offer them entirely. However, it also reflects historic stakeholder concerns about unreliable marking and poorly functioning assessments which are technically distinct from severe grading, but frequently conflated with it when arguments are made that these subjects are 'harder' than others. These include the proportion of native speakers within the cohorts for A level French, German and Spanish and the distorting effect they may have had on the candidate profile; unreasonably high expectations of language learners at A level in the subjects given the available teaching time and lack of other language exposure in the UK; a failure of existing assessments to adequately discriminate at the top end of the performance scale; and a general belief that examiners are inconsistent in their judgements or else mark candidates overly harshly (note that severe marking is distinct from severe grading, although fair awarding should theoretically compensate for harsh marking if carried out correctly and if there are no additional distorting factors at work within the cohort). It is the belief of some stakeholders that these factors will likely become worse within the reformed A level qualifications when they are awarded for the first time in 2018 (although some perceive this as an opportunity to 'reset' student perceptions of the subject), particularly as the content of the new examinations is deliberately more challenging, and there are concerns that severe grading based on an incorrect standard carried forward from the legacy qualifications which is already misaligned will mean that candidates are unfairly penalised.

Some of these factors could be contributing to the apparent difficulty of A level French, German and Spanish under statistical measures of subject difficulty such as Rasch and Comparative Progression Analysis, or otherwise lead to a situation where students do not receive 'fair' grades. However, as they are distinct from the concept of grading severity they are generally issues which would be more correctly and effectively addressed through means other than adjustment to grade boundaries – for instance, through further work to refine assessments. Even in relation to assessment concerns where this is not the case, such as the belief that native speakers who perform better in these subjects than might be expected from prior performance data may be 'stealing' the A and A* grades which should go to other, able students, Ofqual has already undertaken work and made adjustments to grading standards to account for these (although there is a suggestion from stakeholders that we did not go far enough in the case of native speakers).

These issues are discussed in greater detail in the next section of this report. Whilst they may be evidence of legitimate concerns of stakeholders about assessment in A level French, German and Spanish, they are not necessarily to be considered as evidence of grading severity which might lead us to adjust grading standards.

A final argument made by HE subject representatives, similar to that made by subject associations representing teachers and school leadership associations, is that recent political developments will exacerbate the negative impact of severe grading – or else increase the importance of languages to society as a whole to an extent that outweighs the conceptual concerns which have meant that thus far we

are unwilling to make an adjustment on the basis of the statistical evidence alone. In particular, the impact of Brexit is cited as making the undersupply of linguists (and hence A level language study) more acute than before, and as presenting an additional argument for acting to address perceptions of grading severity. Again, these arguments presented alongside the evidence of potential negative trends in these subjects which stakeholders argue contribute to grading severity, and stakeholders feel that holistically they present a compelling case for action:

“Intervention is vital not only for the health of the subject but also for the country at a time when the UK is in danger of becoming inward-looking and increasingly hostile to linguistic diversity and more than ever needs ‘home-grown’ linguists” (Carruthers et al, 2017).

In summary then, the arguments of stakeholders from higher education would ostensibly seem to demonstrate widespread support for an adjustment to grading standards in A level French, German and Spanish (although it should be remembered that they only represent a sample of university language specialists, and are self-selecting to the effect that we have only received the views of those who have felt strongly enough about the subject to write to us), and encompass concerns about historical unfairness in misalignment of standards; negative trends within the subject which are attributed to severe grading and are felt to have the potential to exacerbate issues in the future; concerns about unreliability of assessment; and a belief in the future significance of languages for wider society – and that they feel that together these factors are compelling enough for Ofqual to act, even if we are not satisfied that an adjustment to grading standards alone will be sufficient to address the issues that they have identified: “Change will only materialise if young people are encouraged to feel that languages are as valued as much as other subjects, and they are enabled to pursue the subject with the confidence that their academic efforts will be fairly rewarded” (Carruthers et al, 2017). The implication would seem to be that adjustment of grading standards per se would lessen the perception that these A level subjects are hard and make some contribution to addressing the issues in them, even if we are not able to conceptualise what we are seeking to achieve in making such an adjustment. Certainly, stakeholders in higher education see action by Ofqual as a necessary step in “equipping young people with the intercultural competence and communicative versatility they will increasingly need if they are to play an effective part in a globally oriented world”, and that severe grading is one factor negatively influencing the status of languages which it is in the power of the education system to address – and hence “urge Ofqual to play its crucial part”.

2.4.3 Interaction with other issues

As referenced above, it is important to bear in mind that whilst the concerns of stakeholders about severe grading in this subject are longstanding, Ofqual has taken recent steps to address specific assessment issues with A level modern foreign languages, which had previously manifested in depressed outcomes. In some cases these have been actions which were either in progress at the time that much of the stakeholder representation detailed above was received, or else where we might expect a period of ‘bedding in’ to be required for the impact of these changes to be

fully realised – particularly in terms of the public perception of subject difficulty, which is likely to change relatively slowly even where we announce changes.

Recent intervention taken by Ofqual into the assessment and awarding of A levels in French, German and Spanish includes improvements to assessment design which came into effect in 2015, and which sought to improve the ability of questions and mark schemes to discriminate between high-performing candidates; the re-baselining of A level A* outcomes in 2015 with the steer to exam boards to only use positive awarding tolerances at that grade; the 1% adjustment to take into account the influence of native speakers within languages cohorts which came into effect in awarding in 2017; and the reforms to A levels (including the reweighting of the non-examined assessment and assessment objectives) which were examined for the first time in summer 2018. These actions have generally been received by stakeholders as positive steps to address assessment issues in these subjects.

One recent press release by the Association of School and College Leaders, in response to the adjustment to A level A/A* grade percentages to account for native speakers, stated that in light of other recent work to address the “incorrectly low percentage of candidates gaining an A* grade since its introduction in 2010” that this meant now “that ML candidates are being treated comparably to other ‘facilitating’ arts subjects such as History” (Association of School and College Leaders, 2017), implying that they feel this issue has now been resolved. Given that we have acted on the basis of our research, and stakeholders seem to be satisfied with the outcome, we might reasonably conclude that we can discount native speakers from our consideration of other potential explanations of the ‘severe grading’ we see under statistical measures of difficulty.

However, there are also suggestions from some stakeholders that the steps we have taken have not been adequate to address the issues being experienced in these subjects, and that further action is required. This is particularly the case in regard to German, where HE representatives are concerned that a blanket 1% adjustment at A* in A level French, German and Spanish did not reflect the research which indicated differing proportions of native speakers present within the cohort for this subject: “failure to differentiate between languages in the adjustment will simply cement current anomalies for the future, and exacerbate the problems caused by longstanding unfair grading” (Kohl and Mentchen, 2017). If we feel that there is a continuing impact in certain languages from native speakers then this may lead to us instead making another limited, adjustment to grading standards at A*/A on that basis, rather than seeking to align on the basis of statistical measures of subject difficulty; without setting a precedent for our conceptions of inter-subject comparability. The rationale behind such a decision would need to be clearly explained.

We must be mindful, when proposing adjustments to grading standards on the grounds of negative trends in these subjects, of the extent to which the concerns identified by stakeholders may already have been addressed (either partially or fully) by the work undertaken on A levels in French, German and Spanish – and also whether there is the potential that the apparent severe grading of these subjects, if in fact due to factors other than severe grading, may no longer be observable when we analyse the results of the first reformed qualifications in 2018. If so, adjustment now might risk ‘fixing’ these issues twice.

2.4.4 Proposed changes

Stakeholders from teacher subject associations and school leadership associations, and those representing the subject within higher education, agree that the approach of Ofqual to addressing potential issues of severe grading should be a pragmatic one – both in terms of the basis on which we determine to act (in recognition of negative impacts on the subject which *may* be the result of grading severity rather than on the basis of compelling evidence of causation where we cannot conceptualise it), and in terms of our policy response.

Indeed, in responding to the original policy options proposed as part of our consultation on approaches to inter-subject comparability in 2015, languages stakeholders were clear that whilst taking no action would not be acceptable as it would do nothing to address concerns in the subject, the two arguments that would have seen outcomes statistically reweighted to provide ‘absolute’ comparability were also not felt to be appropriate as they did not reflect the reality of variable candidate ability within the cohort – and that even their favoured option, adjusting grading standards to ensure ‘true’ comparability, was not feasible for all subjects. Instead they claim that subjects could more realistically be divided into semi-cognate bands of statistical difficulty (Myers and Blow, 2017).

Whilst one stakeholder has argued that any adjustment “should be significant and not minor, given that you are keen to make [it] a one-off”, stakeholders generally argue in favour of an adjustment which means that students receive outcomes which allow them a reasonable expectation of performing as well in A level French, German and Spanish as they might in other ‘facilitating’ (i.e. traditionally academic) A level subjects, and believe that this will remove the ‘strategic’ elements of the decisions made by students and schools about subject choice and provision. Whilst they recognise some of the other issues present in languages which are not the result of severe grading, they regard a grade standard adjustment as at least presenting a ‘level playing field’ in which A level languages can compete fairly in the option blocks of students. Furthermore, the subject associations urged us to take action before the first award of the reformed qualifications in 2018.

With this in mind, the subject associations are in favour of an adjustment which brings A level French, German and Spanish into alignment with non-cognate but traditionally ‘facilitating’ humanities subjects such as history or geography. It should be noted that there is a significance in which of these subjects was chosen however, as they do not appear to be equally difficult under Rasch and Comparative Progression Analysis (and in light of the updated Rasch analysis for 2017 results which has seen A level languages shift slightly to the more lenient end of the spectrum, geography and Spanish are now equally ‘difficult’ at certain grades).

Language stakeholders representing teachers and leadership associations are of the view that such an adjustment would represent only a small “one-off step change” to standards (Myers and Blow, 2017), and that the changes required in terms of adjustment to the cumulative percentages achieving a given grade are within the normal variation seen year-on-year in the application of awarding tolerances to exam board grading outcomes, and therefore likely to go relatively unremarked – avoiding any accusations that the subject is being ‘dumbed down’. Such adjustments could arguably be of a similar order to those made at grades A* and A to address the issue

of native speakers⁵, and even the negative impacts of more substantial adjustments could be minimised by making them over a number of years. Whether small changes of this order would serve to wholly address the perception of severe grading however is another matter, as any change which is potentially unnoticeable may not be perceived by all stakeholders as sufficient to address what is seen as (and widely held to be) a persistent issue.

The overwhelming view of stakeholders is that we should be taking action to address severe grading at GCSE, where they feel the effects are far more pervasive and detrimental to the subject, and which they identify as a causal factor in issues with uptake “downstream” at A level. In particular, teacher subject associations are of the view “the fact that modern languages GCSEs are graded more severely than other GCSEs... is a disincentive for young people to study the beyond age 14” (Barton et al, 2017). They also feel that severe grading of modern language GCSEs means that “schools are penalised in the league tables if large numbers of pupils take them”, and that this is dissuading schools from offering these subjects at key stage 4.

Whilst the majority of the teacher subject and leadership associations stakeholders do not go as far as suggesting that action to address apparent severe grading at key stage 4 should take priority over our work on A level (and HE subject representatives would certainly not agree with this view), they have expressed concerns that the scope of our work this year does not also encompass GCSE: “we are however deeply disappointed that the issue of severe grading at GCSE MFL is not being addressed at all... this would have been an ideal opportunity to recognise the issue both at GCSE and A level and create a level playing field for our students” (Barton et al, 2017).

It is certainly the case that issues in uptake at GCSE could be reducing the number of students who are eligible to progress to study the subject at A level, and thus contributing to the decline in numbers studying French, German and Spanish – but they may also be contributing in other ways. The negative impacts identified at A level which are attributed to severe grading (students selecting away from the subject; decline in teaching quality due to a shortage of qualified linguists; impact on student motivation; reduced time available for teaching; decreased opportunities to study more than one language) may well also all be present at GCSE: with the added consideration that they could in turn be contributing to the apparent underperformance of candidates at A level and either giving the impression of, or intensifying the impacts of the actual presence of, severe grading. This potential relationship between severe grading at GCSE and A level is something we will need to keep in mind when reaching a decision on whether to adjust grading standards in these subjects, and in particular the likely impact of a future adjustment to standards at GCSE (were we to make one) on top of what we might decide to do in A level French, German and Spanish.

⁵ Adding 1% to the prediction at A/B in French, German and Science in 2017 as a result of the native speakers research led to an increase of 1.8%, 1.4% and 2.6% at A and 1.8%, 0.5% and 2.6% at A* respectively.

2.5 Awarding Organisations' views

2.5.1 Awarders' views

As those ultimately responsible for maintaining grading standards in these qualifications, judgemental evidence was gathered from the A level awarding panels for these subjects via questionnaires distributed on behalf of Ofqual by the exam boards. These were completed at or shortly after the awarding meetings. Awarders were asked to reflect on the standards set in their subject in the summer 2018 series, and whether in their professional opinion it was correct. They were also asked to consider whether they felt standards were appropriate relative to other A level subjects, how acceptable an adjustment to grading standards might be in terms of the impact on candidate performance standards at the judgementally set grade thresholds of A and E, and whether the current grading standard might be having any negative impacts.

The responses received have been summarised below by subject.

French

Two of the four boards did not provide views from awarders on grading standards in this subject, including one board who have now ceased to offer A level languages. Of the two boards that did, the awarding panels at both were divided over whether the standard was correct or potentially slightly severe. At one board the panel agreed that there was scope to relax the standard marginally at both the A and E thresholds, although this seemed to be primarily due to concerns about the status of French as a 'difficult' subject rather than in response to the quality of candidate work. At the other board, awarders suggested that any change should be at grade A only – although one dissented, arguing that the standard at this threshold should in fact be raised as they expected candidates obtaining an A to demonstrate a proficiency in the languages which exceeds what is currently being demonstrated. Views on the necessity of lowering standards were again largely couched in terms of the experience of teachers, with the justification offered for relaxing standards at grade A the argument that schools "are finding it very hard to get decent grades for decent pupils under the gradings as they are".

Neither panel expressed a clear view on whether standards in French were comparable to those in other subjects, although in both cases awarders made the point that they felt that the cognitive demands in languages were inherently more challenging than in other subjects. Both panels recognised the perception amongst teachers and students that this subject is more difficult than other A levels, and agreed this was likely discouraging students from studying the subject and contributing to the decline in entries. However, neither panel could agree if this perception was justified – with one awarder arguing that this 'difficulty' was more to do with inadequate resourcing and teaching time at key stages 3 and 4. Both panels felt that such an adjustment to standards would be acceptable amongst stakeholders, but awarders at one board disagreed over whether an adjustment to standards alone would be sufficient to address the decline in entries in this subject. The same panel also felt that perceptions of difficulty may wane in coming years due to the reformed GCSEs providing better preparation for A level study, and warned

that lowering standards at grade A would likely make it more challenging for universities to identify students possessing the necessary grammatical knowledge to cope with further study.

German

Two of the four boards did not provide views from awarders on grading standards in this subject, including one board who have now ceased to offer A level languages. Of the two boards that did supply views from their awarding panels, awarders from one felt that grading standards were broadly correct whilst the other panel felt that the subject was graded severely at A. Where awarders felt that the standard required of candidates was incorrect, this was attributed to changes in the assessment made as part of the reforms rather than a longstanding misalignment of the standard.

Awarders from one of the panels were either satisfied that the standard of A level German was suitable in comparison to other subjects, or did not feel able to make a judgement. The other awarding panel however felt strongly that German was harder than other A level subjects at grades A and A*, something which they attributed to the continuing impact of native speakers within the cohort preventing non-native speakers from accessing the highest grades. This informed their argument to reduce the standard at grade A. Awarders from the other board were of the view that the standard should be maintained. Both panels felt that there was a widespread perception that German is more difficult than other A levels, and that this was dissuading students from studying it in the belief that they will secure better outcomes elsewhere. However, only at one board did awarders feel that this viewpoint was justified. Awarders on the other panel were unsure whether this perception was representative of genuine grading severity, or if German is inherently more demanding than other subjects (and possibly even other languages, which they felt lacked the same syntactical complexity). Similarly, whilst one panel was confident that an adjustment at grade A to address the issue of native speakers would be sufficient to begin to address issues around post-16 language uptake, awarders at the other board felt that such intervention may be unnecessary as the reforms to GCSE will better prepare candidates for progression to A level and their experience of the subject will improve. Whilst both awarding panels generally felt an adjustment to standards might help to address declining entries in the short term, this was not a unanimous view and one awarder felt that negative perceptions of the subject are likely to persist regardless.

Spanish

Three out of four of the exam boards provided views from their awarders on standards in the subject. In the case of the board who did not, this was because they have now ceased to offer A level languages. The awarding panel from one board were of the view that the current grading standard in A level Spanish is correct (with students obtaining the 'right' grade in this subject), awarders at another board felt that the subject was graded severely at grade A, and the final panel were divided over whether the standard was correct (particularly now that an adjustment had been made to take into account the impact of native speakers) or slightly severe. All three panels felt the standard at grade E was correct, differing only in their view of whether an adjustment would be acceptable at A.

Two of the three panels felt that Spanish was graded more severely than other A level subjects, believing that students were likely to achieve their lowest grade in the

subject and arguing for greater comparability with other, non-language A levels. They also expressed the view that the cognitive demand upon students studying modern foreign languages was greater than in other subjects. All of the panels felt that Spanish was recognised as being more difficult than other subjects by students, particularly in terms of terms of achieving grades A* and A, and that this was leading them to study alternative A levels which they regarded as 'easier'. This was attributed to the impact of native speakers, which awarding panels felt the evidence of speaking tests in particular indicated had not been adequately addressed. Awarders on one panel expressed concerns that this effect would be exacerbated by the decoupling of AS and A level, and the general move to studying only three subject post-16. All three panels felt that relaxing standards at grade A would help to address this decline in uptake, but two of the panels caveated this by asserting that it was also important to ensure that there was meaningful progression from GCSE to A level so that students were adequately prepared for further study and that universities continued to value the qualification for admissions purposes.

2.5.2 Viewpoints of the Awarding Organisations

In addition to gathering judgemental evidence on potential grading severity and the appropriateness of an adjustment to grading standards from their awarding panels, three of the four exam boards also offered their own analysis of the evidence presented in these reports. This is summarised below.

Statistical Evidence

Those responding recognised that the statistical evidence presented in the report seemed to suggest that A levels in French, German and Spanish are more difficult than the average A level. However the boards also noted that the report identified some significant limitations of the various forms of statistical analysis which have been employed, which they feel present significant conceptual challenges to drawing conclusions about 'difficulty', and the advisability of adjusting grading standards.

Noting that the statistical measures of subject difficulty share the assumption that all subjects are measures of a single latent trait (i.e. a 'general academic aptitude'), the boards expressed doubts both about the unidimensionality of what is purported to be being tested in different A level subjects (whether we really can expect that a given student will perform just as well in history as they do in mathematics), and what the shared latent trait being measured actually represents. Thus whilst the statistical evidence may indicate that there are differences between A level subjects, the evidence does not provide an explanation of what these differences are – they could be the result of more severe grading, but they may also reflect differences in the resources which schools and government dedicate to particular subjects, more or less demanding subject content, or variations in overall teaching quality and effectiveness. It may also reflect differences in the 'step up' from GCSE to A level between different subjects.

Another limitation to the statistical evidence recognised by the exam boards is the extent to which the statistical models may be compromised by either non-representative samples or non-random missing data – evidenced by the simulations conducted by Bramley (2016) – which may be contributing to biased estimates of subject difficulty. This is particularly an issue given the relatively unrestricted nature of students' subject choices at A level. Rasch models assume that subjects are chosen at random, and that candidates in the years considered selected only those subjects in which they would perform well. However, the boards are not convinced that this is the case. One board suggested that the possibility that students may choose certain A level subjects because they are deemed to be 'facilitating' (i.e. regarded as providing greater opportunities for further study or employability) even where it is not one of their strongest subjects makes facilitating subjects particularly problematic under this kind of analysis. All of the subjects considered in this report are so-called 'facilitating subjects'.

Boards also felt that there were differences in the expectations of prior learning between A level languages and some of the other subjects considered within the statistical analysis of subject difficulty. In particular the fact that students must previously have studied a GCSE language (often to a minimum standard of competency) to progress to study it at A level, whereas some subjects students will be studying for the first time, was viewed as potentially being a reason that the 'difficulty' of different A levels may justifiably differ.

One board also made the point that whilst the report considered the evidence from the perspective of addressing potential grading severity by considering whether to lower grade boundaries in languages, it did not give any consideration to whether standards in other, apparently lenient subjects should be increased. However, such a response to inter-subject comparability was specifically excluded from the scope of this research as a result of the Ofqual consultation in 2016.

In light of the potential limitations outlined above, the boards shared the view expressed by Benton and Bramley (2017) that these limitations means that is not possible to align subjects solely on the basis of statistical analysis.

Impact of Reform

The boards also stressed that the statistical evidence presented in the technical reports for both A level science and languages was based upon analysis of results data from the legacy A levels sat in 2013 and 2017, rather than the reformed qualifications. Whilst they acknowledged that the Comparable Outcomes approach that Ofqual adopt to awarding means that grading standards from the legacy qualification have been carried forward and will continue to be maintained in the new A levels (along with any historical severity in grading which may have existed in

these subjects previously - particularly since in 2017 and 2018 outcomes were closely aligned with predictions), there have been substantial changes to the design of these assessments which may arguably have resulted in a change in content standards.

In particular, one board noted that as the reformed A levels in modern foreign languages were awarded for the first time in summer 2018, it would take time and further research to fully understand what the impact of those changes might be. This is complicated further by factors such as the saw-tooth effect, which may compound the apparent difficulty of these subjects in the first few years of awarding as teachers adjust to delivering the new qualifications, with candidate performance progressively improving as they do. For this reason the board were of the view that any adjustment should not take place immediately, but after a period in which the new qualifications were allowed to bed in.

Similarly, it was also noted that it was still unclear what the impact of the decoupling of AS and A level might be upon the apparent 'difficulty' of these subjects, something which may not be evident until all of the reformed A levels have been introduced and assessed. Whereas previously students routinely began studying four AS levels (likely choosing to discontinue one subject after a year) there has now been a widespread movement in schools to offering three A levels only. Whilst this has been identified elsewhere in the report as providing a potential alternative explanation for the decline in uptake in certain subjects (A level entries being a 'zero sum game'), one board have hypothesised that this will also reduce the number of students taking subjects across disciplines (e.g. a combination of sciences and humanities) and lead to an increased propensity to focus A levels choices within the same subject group. They suggest that this may have an impact on the relative difficulty of different A level subjects under statistical measures which are predicated on the assumption of a generic linking construct such as Rasch and Comparative Progression Analysis, and recommend that further research is carried out to establish the impact of AS and A level decoupling – rather than attempting to adjust grading standards on the basis of data from the legacy qualifications.

Several of the boards also made the point that AS and A level specifications must build on the skills, knowledge and understanding proscribed at GCSE, and that there is likely to be an interaction between how effectively GCSEs prepare students for A level study and perceptions of subject difficulty at key stage 5. This interaction is likely to have changed following the reforms to GCSE, which explicitly set out to increase the content demands of these subjects (which was not the case at A level). The reformed GCSEs in science and modern foreign languages were awarded for the first time in summer 2018, and the first students progressing to A level from the reformed qualifications will take their assessments in summer 2020. Similarly, one board emphasised that policy changes associated with the reforms are likely to have an impact on school behaviours (and thus the experiences of students) which may

result in changes to the apparent 'difficulty' of these qualifications (for instance, the change in the number of students taking combined science as opposed to three separate sciences at GCSE). Again the impact of these changes will not become clear until 2020, and one board argued that any action on grading standards should be postponed until this time.

Judgemental Evidence

The boards gathered individual feedback from members of their awarding panels via questionnaires, which have been summarised above. However, the boards also provided their own analysis of the judgemental evidence they submitted.

The boards noted that the views from awarders often differed, and that on the majority of panels there was some diversity of opinion as to whether, in their professional judgement, an adjustment to grading standards would be acceptable. Overall they reported that where awarders were in agreement this was generally that the current standard was correctly set. Where awarder opinion was divided, as it frequently was amongst the modern foreign language panels, it was not always the case that the disagreement was over whether the standard was correct or too severe – some feeling that the standards were in fact too lenient.

It was stressed that the small cohort for A level German (in comparison to sciences, and to French and Spanish where entries are comparable) meant that awarders were making judgements on the basis of more limited evidence than other subjects.

On balance, the boards felt that overall the judgemental evidence did not provide a compelling argument that grade boundaries at A level should be adjusted, either because awarders were overwhelmingly of the view that standards were correct or because opinion was divided to the point that the evidence was inclusive. Boards did report however that the majority of awarders recognised that stakeholders perceived these subjects as being hard (rightly or wrongly), although one board pointed out this was perhaps unsurprising given the extent to which their panels are invested in the current standard. Awarders felt that these perceptions may be contributing to the decline in uptake in modern foreign languages as candidates regard it as easier to obtain a high grade in other subjects, reflecting the view of stakeholders within the subject community.

The boards reported that their awarders generally felt that an adjustment to grading standards, particularly at grade A, might help to address issues of declining uptake in A level languages. On the other hand, they did not feel that there was any compelling evidence that an adjustment would lead to any increase in entries for science subjects – which they noted had already seen substantial rises in recent years, likely as a result of recent initiatives to promote STEM subjects.

Boards also cited the research conducted by Ofqual and the DfE into students' subject choices, which indicate that the motivations behind study decisions are complex and that utility and enjoyment are often greater considerations than difficulty. One board did consider that enjoyment may be linked to some extent to perceived success in the subject, and that this may exacerbate the impact of beliefs about subject difficulty on uptake. In regard to the influences on students' choice of whether to study A level languages, the complexity of the issue was highlighted by reference to research by Macaro (2008) which concluded that the decline in language learning (and to some degree the perception they are difficult) was the result of educational policy decisions and the lasting impact they have had on language pedagogy and pupil experience at key stages 3 and 4. This also illustrates the point that the factors influencing student decisions are liable to change over time.

Potential Risks

The exam boards were mindful that any re-evaluation of the standard should weigh the impact of the current standard against the likely impact of an adjustment, and in particular that this should be considered in the context of the purpose of A levels.

Two purposes in particular were regarded as being at risk from an adjustment to standards, these being the requirements of A levels:

“To define and assess achievement of the knowledge, skills and understanding... needed by students planning to progress to undergraduate study at a UK higher education establishment, particularly (although not only) in the same subject area.”

“To permit UK universities to accurately identify the level of attainment of students.”

Here it was felt that there was a tension between the purpose of A levels to facilitate progression in a particular subject, and the purpose to facilitate progression to higher education in general. Boards were concerned that any change to the standards at A level, particularly at grade A, might result in a loss of discrimination within subjects at the top end and lead universities to respond by raising their entry requirements within that subject. However, it was also acknowledged that outside of a particular subject A levels are presented as having equal currency according to their UCAS point tariff (although admissions tutors may have their own views on the relative value of a qualification), and that this means that some students may be disadvantaged if they have taken a 'hard' subject when applying to university, compared to those who have taken an 'easy' one. Boards recognised that any change in the value which is assigned to different A levels was beyond the remit of Ofqual, and we can only seek to assure comparability of subjects through changes to grading standards. One possible solution suggested was to adjust the standard at grade E (which would have a knock-on effect on the standard of the intermediate grades) whilst maintaining the standard at grades A* and A to support progression in the subject. However it was acknowledged that this was unlikely to address the perception of difficulty in these subjects.

Increases in university entry requirements were felt to be a particular risk if changes were to be made mid-cycle rather than at the defined end-point of a specification, as universities would have to adjust their understanding of the performance standard represented by passing and higher grades to reflect the change in the difficulty of the

qualification. Concerns were expressed that either some universities may not fully grasp the new standard, unfairly penalising some students and advantaging others, or that students in adjacent cohorts would be disadvantaged – either those in the cohort immediately preceding any adjustment who would be competing for jobs and university places with students who achieved the same grade for a lower level of performance, or students in successive cohorts who were regarded as holding a devalued qualification. To mitigate this risk, boards felt that any adjustment should be carried out incrementally.

Ultimately, boards were of the view that for greater comparability to be achieved between subjects, it may be necessary to prioritise one of the stated purposes of A level qualifications over the others. Exam boards argued in favour of preserving the current standard to enable universities to continue to effectively differentiate between candidates.

Conclusions

Overall exam boards were supportive of re-evaluating grading standards in A level modern foreign languages to ensure that they remain appropriate, but did not necessarily feel that either the statistical or judgemental evidence presenting a compelling case for an adjustment – or a clear indication of what a more appropriate alignment of subject difficulty might be.

However boards did feel that this was an issue which warranted further consideration, and in particular that it was reasonable to keep grading standards under review generally because the nature of comparable outcomes (which maintains the relationship between prior attainment and results in a given subject year-on-year) means that any contextual factors which may have contributed to an increased or decreased level of difficulty in a specific subject when comparable outcomes was first introduced have been reflected in the grading standard from that point forward, regardless of whether those factors have subsequently changed.

It was also noted that difference in difficulty between subjects was likely also present at GCSE, and that if adjustment were to be made at A level it would also be necessary to consider subject alignment lower down. In doing so, there would likely be similar tensions between prioritising parity and ensuring appropriate progression from GCSE to A level study in a given subject.

Finally one exam board advocated that if any adjustment were to be made to grading standards this should not occur immediately, but as part of any future reform to A levels.

2.6 Higher Education Representatives' Perception of Grade Standard Adjustment

Curcin, Black and He (2018) investigated how a potential grade standard adjustment in A level French, German, Spanish, physics, chemistry and biology might be perceived by higher education (HE) representatives, as the key users of A level qualifications. This study investigated whether an adjustment, and what magnitude of adjustment in each subject: would be discernible to the participants; would be

acceptable to them; and might impact on the utility of those grades for admissions purposes.

To identify the maximum possible adjustment to each grade boundary in each subject, the Rasch-based method described in He and Stockford (2015, also see next section) was used to identify the range of subject difficulty on a single scale, and then identify the implied grade boundary adjustments to bring sciences in line with mathematics, and languages in line with geography. Determining the subjects to align each cognate group to was essentially an arbitrary decision, but one which was necessary for the study and which was judged to be reasonable in light of the similarity of the balance between skills and knowledge in these subjects to those considered in the study; their relative difficulty under Rasch analysis; and the views of stakeholders. In particular, the undergraduate courses which students might use these subjects to access (and which may be regarded as comparable for admissions purposes by HE) were considered, as well as feedback which indicates that stakeholders within mathematics and geography are satisfied with A level standards.

For each subject, representatives from a range of HE institutions as well as learned body representatives and Ofqual subject experts took part in panels which were asked to review samples of work from A level examinations (student scripts) at and below each of the A*/A, A/B, B/C and C/D grade boundaries, including the scripts on each indicative adjusted grade boundary as suggested by Rasch analysis. The scripts that represented the actual qualification grade boundaries were treated as 'benchmark' scripts, i.e. scripts representing the current grade boundary standard.

The scripts in the sample represented the whole work of candidates that were examined at the end of the qualification. For science subjects, for each script/student, this comprised the totality of each student's A level examined work. For the languages this included paper examinations as well as the speaking examination (audio file) but only the examined work taken at A2 level.

For each script in each grade boundary script set, the participants were asked to answer either:

- Whether the student who wrote it was equally deserving of admission to their institution as the ones who wrote the 'benchmark' scripts on the current grade boundary (this decision was only made by those participants whose institutions use the higher grade as an admissions criterion), or
- whether it was similar, better or worse than the 'benchmark' scripts

After this exercise, the participants discussed grade standards issues in panels.

While the Rasch analysis suggested much more of a 'gap' in alignment between the sciences and mathematics than between the languages and geography, the script

review exercise suggested that there may be less appetite for grade standard adjustment in the sciences, in contrast with the languages.

Furthermore, more often than not in the science subjects the discussion by the panel indicated a lack of acceptance of any adjustment, perhaps surprisingly even on those few occasions where the collated experts' individual judgements indicated that some level of adjustment might be acceptable. 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. This was also the case even where the statistical adjustment was not proposed at all, as was the case for some boundaries in Spanish.

In summary, this work provides some evidence from the script review of the acceptability of limited potential grade boundary adjustments in some subjects. This is more the case for modern foreign languages than the sciences. However, where this was the case in the sciences, the discussion by the panels did not always support making an adjustment. Detailed results are discussed in the report annexed to this document.

3 Impact of statistical alignment of standards on outcomes

If standards were to be aligned for different subjects, based on inter-subject comparability studies using statistical methods, the grade boundaries for some of the examinations would need to be changed. As grade boundaries can be viewed as the operationalization of performance standards, any change in grade boundaries would imply a qualitative change in performance standards from those established for some of the examinations (e.g. subjects which are either too "easy" or too "hard" based on statistical comparisons). The distributions of grades will also change accordingly. Some of the distributions may become less effective in differentiating the candidates (for example, if the resultant grade distribution becomes highly skewed towards the top grades or bottom grades). This would have an impact on the interpretation of grades. This section looks at the impact of statistically aligning standards between subjects on the quality of candidate performance at adjusted grade boundaries and on grade distributions for some of the subjects under consideration.

3.1 Change in grade boundaries and grade distributions

3.1.1 Aligning to the mean of all subjects (based on Rasch analysis)

Table 10 shows the original grade distributions for A level French, German and Spanish from the 2017 examination series. All three qualifications are unitised⁶ and use uniform mark scale (UMS)⁷. The effect on A* was not examined due to the non-linear rules for determining those candidates that achieve an A* in specifications using UMS. Instead, the impact on the combination of A* and A was examined. Percentage changes in candidates at individual grades and the changes in the cumulative percentages as well as the original and adjusted qualification level UMS grade boundaries (GBs), after alignment of standards statistically to the average of all subjects based on Rasch analysis, are also shown in the table. The alignment involved shifting the original grade boundary mark at a specific grade by the proportion of grade width determined by its relative difficulty (see He and Stockford, 2015, for details of the boundary mark adjustment process). For French, the cumulative percentage of candidates at grade A would go up by about 3% (from slightly over 39% to over 42%). At grades B and C, the cumulative percentage of candidates would increase by slightly over 5%, and the UMS boundary mark would need to decrease from 280 to 270 at B and from 240 to 224 at C. The changes in grade boundaries suggest that the performance standards at the adjusted grade boundaries would be lower than the original performance standards at these grades. For German, the cumulative percentage of candidates would increase by about 3% at A and slightly less than 3% at B and C. For Spanish, the change in cumulative percentages across the grades would be negligible in the higher grades.

Table 10 Changes in percentages of students receiving individual grades and cumulative percentages for A level French, German and Spanish in the 2017 examination series after alignment of standards statistically to the mean of all subjects (based on Rasch analysis).

Subject (N)	Grade distribution (%) and grade boundary change (% of max UMS marks)						
		A*+A	B	C	D	E	U
French (8450)	Original (Ind.)	39.36	26.51	18.69	10.71	3.80	0.93
	New (Ind.)	42.27	29.03	18.75	7.05	2.45	0.45
	Change (Ind.)	2.91	2.52	0.06	-3.66	-1.35	-0.49
	Original (Cum.)	39.36	65.87	84.56	95.27	99.07	100.00
	New (Cum.)	42.27	71.30	90.05	97.10	99.55	100.00
	Change (Cum.)	2.91	5.43	5.49	1.83	0.49	0.00
	Original UMS GB	320	280	240	200	160	
New UMS GB	316	270	224	186	142		
German (3387)	Original (Ind.)	41.10	28.14	17.95	8.74	3.34	0.74
	New (Ind.)	44.52	27.40	17.95	7.85	2.10	0.18
	Change (Ind.)	3.42	-0.74	0.00	-0.89	-1.24	-0.56
	Original (Cum.)	41.10	69.24	87.19	95.93	99.26	100.00

⁶ Unitised or modular examinations or qualifications used in England: Assessment units in the examination or qualification can be taken throughout the course of study.

⁷ UMS mark is a scaled score used to ensure the comparability of raw marks from different examination series. See <http://www.aqa.org.uk/exams-administration/about-results/uniform-mark-scale> for a detailed explanation.

	New (Cum.)	44.52	71.92	89.87	97.73	99.82	100.00
	Change (Cum.)	3.42	2.69	2.69	1.80	0.56	0.00
	Original UMS GB	320	280	240	200	160	
	New UMS GB	316	275	230	186	142	
Spanish (7718)	Original (Ind.)	37.17	30.28	19.11	9.10	3.49	0.86
	New (Ind.)	37.17	30.28	21.15	8.58	2.29	0.53
	Change (Ind.)	0.00	0.00	2.03	-0.52	-1.19	-0.32
	Original (Cum.)	37.17	67.45	86.56	95.66	99.14	100.00
	New (Cum.)	37.17	67.45	88.60	97.18	99.47	100.00
	Change (Cum.)	0.00	0.00	0.00	2.03	1.52	0.32
	Original UMS GB	320	280	240	200	160	
	New UMS GB	320	280	234	190	147	

3.1.2 Aligning to A level Geography (based on Rasch analysis)

Table 11 shows the original and expected grade distributions, changes in percentages of candidates at individual grades and in the cumulative percentages, the original and adjusted qualification level UMS grade boundaries, after alignment of their standards to those of geography based on Rasch analysis for A level French, German and Spanish from the 2017 examination series. These changes are only very slightly different from those shown in Table 8. Changes in grade distributions are negligible for Spanish. For French and German, the changes in cumulative percentages of candidates are less than 4% for most of the grades.

Table 11 Changes in percentages of students receiving individual grades and cumulative percentages and grade boundaries for A level French, German and Spanish in the 2017 examination series after alignment of standards statistically to those of Geography (based on Rasch analysis).

Subject (N)	Grade distribution (%) and grade boundary change (% of max UMS marks)						
		A+A	B	C	D	E	U
French (8450)	Original (Ind.)	39.36	26.51	18.69	10.71	3.80	0.93
	New (Ind.)	42.84	27.25	19.08	7.63	2.64	0.56
	Change (Ind.)	3.48	0.75	0.39	-3.08	-1.16	-0.38
	Original (Cum.)	39.36	65.87	84.56	95.27	99.07	100.00
	New (Cum.)	42.84	70.09	89.17	96.80	99.44	100.00
	Change (Cum.)	3.48	4.22	4.62	1.54	0.38	0.00
	Original UMS GB	320	280	240	200	160	
	New UMS GB	315	272	227	188	148	
German (3387)	Original (Ind.)	41.10	28.14	17.95	8.74	3.34	0.74
	New (Ind.)	44.52	26.90	17.98	8.09	2.24	0.27
	Change (Ind.)	3.42	-1.24	0.03	-0.65	-1.09	-0.47
	Original (Cum.)	41.10	69.24	87.19	95.93	99.26	100.00
	New (Cum.)	44.52	71.42	89.40	97.49	99.73	100.00
	Change (Cum.)	3.42	2.18	2.21	1.56	0.47	0.00
	Original UMS GB	320	280	240	200	160	
	New UMS GB	316	276	232	189	147	
Spanish (7718)	Original (Ind.)	37.17	30.28	19.11	9.10	3.49	0.86
	New (Ind.)	37.17	29.55	21.04	8.91	2.69	0.62
	Change (Ind.)	0.00	-0.73	1.93	-0.18	-0.79	-0.23
	Original (Cum.)	37.17	67.45	86.56	95.66	99.14	100.00
	New (Cum.)	37.17	66.73	87.77	96.68	99.38	100.00

	Change (Cum.)	0.00	-0.73	1.20	1.02	0.23	0.00
	Original UMS GB	320	280	240	200	160	
	New UMS GB	320	281	237	193	152	

3.1.3 Alignment of standards based on CPA analysis

To expand the work on CPA reported by Newton et al. (2017), Benton and Bramley (2017) looked at the relationship between achievement at A level and attainment at GCSE in the same subject for all A level examinations administered in 2014 using data extracted from the National Pupil Database. They then investigated the impact of aligning within subject progression from GCSE to A level to the overall progression of all subjects for A level Mathematics, Biology, Physics, French and History. Table 12 below shows the original cumulative grade distributions and the expected (adjusted) grade distributions for A level French and History. For French, the cumulative percentage at grade A*, A and B would have to increase by about 8%, 5% and 4% respectively. For History, the cumulative percentage at B and C would need to decrease by about 5% and 6% respectively.

Table 12 Comparison of expected cumulative grade distribution after aligning progression from GCSE to A level to the overall progression of all subjects for A level French and History in 2014 (Based on Benton and Bramley, 2017).

A level grade	French		History	
	Expected	Actual	Expected	Actual
A*	14.1%	5.6%	7.8%	6.9%
A	41.5%	36.4%	25.7%	26.7%
B	69.7%	65.7%	52.4%	57.8%
C	87.8%	86.0%	77.5%	83.6%
D	96.3%	96.3%	92.6%	96.1%
E	99.3%	99.5%	98.7%	99.5%
U	100.00%	100.00%	100.00%	100.00%

3.1.4 Impact of small changes in outcomes

To see the impact on grade boundaries of small changes in grade outcomes, Tables 13-16 show the entry size, maximum mark and grade boundaries at A*/A, A/B and B/C for A level French, German and Spanish specifications that had entries greater than 500 in 2018 and the changes in grade boundaries at A*/A, A/B and B/C if outcomes at A* and A were to increase by 0.5%, 1.0% and 1.5% respectively.

Specifications have been anonymised in these tables. An increase of 1.0% at A* and A would require boundary marks being reduced by 1-4 marks at both A*/A and A/B.

Table 13 Entry size, maximum mark and grade boundaries at A*/A, A/B and B/C for A level French, German and Spanish specifications with entries greater than 500 in 2018.

Specification	Number of candidates	Max Mark	Original grade boundary mark		
			A*/A	A/B	B/C
FRENCH A	5427	400	354	310	269

FRENCH B	1251	400	319	271	233
FRENCH C	796	200	168	147	125
GERMAN A	1771	400	345	281	236
GERMAN B	537	400	335	279	234
SPANISH A	5117	400	345	307	261
SPANISH B	1292	400	328	293	247
SPANISH C	689	200	179	157	131

Table 14 Changes in grade boundaries at A*/A, A/B and B/C if outcomes at A* and A were to increase by 0.5% for A level French, German and Spanish in 2018.

Specification	Max Mark	Boundary shift		
		A*/A	A/B	B/C
FRENCH A	400	-1	-1	-2
FRENCH B	400	-3	-2	-1
FRENCH C	200	-1	-1	-1
GERMAN A	400	-2	-2	-2
GERMAN B	400	-1	-2	-3
SPANISH A	400	-1	-2	-2
SPANISH B	400	-1	-2	-1
SPANISH C	200	0	-1	-1

Table 15 Changes in grade boundaries at A*/A, A/B and B/C if outcomes at A* and A were to increase by 1.0% for A level French, German and Spanish in 2018.

Specification	Max Mark	Boundary shift		
		A*/A	A/B	B/C
FRENCH A	400	-2	-2	-3
FRENCH B	400	-4	-4	-2
FRENCH C	200	-1	-1	-1
GERMAN A	400	-3	-3	-4
GERMAN B	400	-3	-4	-5
SPANISH A	400	-2	-3	-3
SPANISH B	400	-2	-4	-3
SPANISH C	200	-1	-2	-1

Table 16 Changes in grade boundaries at A*/A, A/B and B/C if outcomes at A* and A were to increase by 1.5% for A level French, German and Spanish in 2018.

Specification	Max Mark	Boundary shift		
		A*/A	A/B	B/C
FRENCH A	400	-3	-4	-5
FRENCH B	400	-5	-6	-4
FRENCH C	200	-2	-2	-2
GERMAN A	400	-4	-5	-6
GERMAN B	400	-4	-5	-7

SPANISH A	400	-2	-4	-5
SPANISH B	400	-3	-5	-5
SPANISH C	200	-1	-3	-3

3.2 Change in performance standards as judged by experts

Jones (2004) studied the impact of aligning standards statistically on the performance standards of A level German and other subjects using subject pairs analysis (SPA) and expert judgement. The study compared 10 large entry specifications provided by the Assessment and Qualifications Alliance (AQA). A level German was found to be about 0.37 of a grade harder than the average of the 10 specifications investigated. Aligning standards at subject level would require lowering boundary marks at the two judgemental grades (A and E) by 25 UMS marks which represents 5/12 of the subject level grade width (which is 60 UMS marks – 10% of the maximum UMS marks of 600). To examine the impact of new grade boundaries on the quality of performance at the judgemental grades, the adjusted subject level UMS marks were then divided between the contributing units to estimate the adjusted unit level raw boundary marks. Scripts near the adjusted subject level boundaries of grades “A” and “E” were drawn and reviewed by the principal examiners (PEs) for their performance standards by indicating their likely position on a notional grade profile. It was found that none of the “A” scripts were identified as having the standard at A. The majority of them were assigned to the notional grade B. That is, statistical alignment of standards between the subjects would result in substantial decrease in performance standards at the adjusted grade boundaries for German.

3.3 Summary

It has been shown that aligning standards statistically between subjects can have substantial impact on the performance standards and grade distributions. For A level French, German and Spanish, particularly French and German, the quality of performance expected of students, for each grade, would need to be relaxed substantially if they were to be aligned with other subjects statistically.

4. Decision

The document detailing our decision on inter-subject comparability in A level physics, chemistry and biology [can be found here](#).

5. References

- Allen, T. (2016). What sort of selection stops students studying physics. Institute of Physics Blog: <http://www.iopblog.org/role-of-selection/>
- 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-43AE-8DD482A055D8AD1A.html>
- Barton, G., Koglbauer, R., Davidson, S. and Richardson, W. (2017). Letter to Ofqual. Available on line at: <http://www.all-london.org.uk/site/wp-content/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/MFL-Pedagogy-Review-Report-2.pdf>
- Benton, T., and Bramley, T. (2017). Some thoughts on the ‘Comparative Progression Analysis’ method for investigating inter-subject comparability. Cambridge Assessment Research Report. Cambridge Assessment: Cambridge.
- 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.
- 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.
- Cuff, B. (2017). Perceptions of subject difficulty and subject choices: Are the two linked, and if so, how? 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.
- Cutler, M., McLeod, L., Morgan, N., Tracy, C. and Mist, R. (2017). Open letter to Ofqual from Association for Science Education, Institute of Physics, The Royal Society, Royal Society of Biology and Royal Society of Chemistry.
- Dearing, R. (2007). Languages Review. Department for Education and Skills: London.
- He, Q., and Stockford, I. (2015). Inter-Subject Comparability of Exam Standards in GCSE and A Level. Ofqual: Coventry.
- Institute of Physics (2018). Why not physics? A snapshot of girls' uptake at A-level. Available on line at: http://www.iop.org/publications/iop/2018/file_71495.pdf
- 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.
- Lockyer, C., and Newton, P. E. (2015). Inter-subject comparability: a review of the technical literature. Ofqual: Coventry.
- 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.

- Newton, P.E., He, Q., and Black, B. (2017). Progression from GCSE to A level: Comparative Progression Analysis as a new approach to investigating inter-subject comparability. Ofqual: Coventry.
- Office of Qualifications and Examinations Regulation (Ofqual) (2015). GCE qualification level conditions and requirements. Ofqual: Coventry.
- 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.](#)
- Qualifications and Curriculum Authority (QCA). (2008). Inter-subject Comparability Studies. Qualifications and Curriculum Authority: London.
- Stevenson, C. Britton, P., McLeish, T., Pritchard, J. and Price, G. (2016). Open letter to Ofqual from Association for Science Education, Institute of Physics, The Royal Society, Royal Society of Biology and Royal Society of Chemistry.
- Stoet, G. and Geary, D. (2018). The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. *Psychological Science* 29, 581-593.
- 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
- 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
- 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 2018

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:



Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

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