



Qualifications and  
Curriculum Authority

---

# **Grade standards in GCSE modern foreign languages**

---

February 2008

QCA/08/3570

# Contents

Background.....	3
Examination results .....	3
Attainment-based comparability.....	4
Aptitude-based comparability.....	5
Discussion .....	6
Conclusion.....	8

## Background

1. Lord Dearing's Languages Review 'found strong confirmation of the view that the award of grades (in GCSE languages) is more demanding than for most other subjects. This needs to be resolved one way or the other by a definitive study, followed by publication of the conclusions, because the present widely held perception in schools, whether right or wrong, is adversely affecting the continued study of languages through to the GCSE.'
2. The Association for Language Learning (ALL), the main subject association for languages, has prepared a detailed statistical analysis to back their proposal that there should be adjustments to the grading of GCSEs to redress the imbalance between languages and other subjects. As well as presenting this evidence to Lord Dearing and DCSF officials, ALL had three meetings with QCA to present and discuss their analysis. One of these meetings included AQA, Edexcel and OCR.
3. We have considered carefully the arguments ALL put to Lord Dearing but we believe that there are strong counter-arguments and that the action proposed could have harmful effects. The paragraphs which follow explain why we take this view. They are based on the purposes and intentions of the GCSE and A level examination systems in England.

## Examination results

4. One of the concerns expressed to us by schools is that as the number of 16 year olds taking GCSE languages has fallen over the years, the awarding bodies have not compensated sufficiently when awarding higher grades, particularly grade C.
5. We have heard from many schools that their GCSE language results have suffered in recent years. Data from the Department for Children, Schools and Families shows results for the 123 schools (excluding grammar schools) that have maintained a high rate of GCSE languages take-up (over 90 per cent entered in each year from 2003 to 2006). Individual school results do show marked changes from one year to the next. However, 61 schools saw an increase in the percentage of GCSE languages candidates achieving grades A\*–C between 2003 and 2006, while 56 saw a decrease.

Of the 45 specialist language colleges included in this data, 24 saw an increase and 19 a decrease.

6. Using data from all maintained schools it is possible to split students into groups with average key stage 3 test levels of 4, 5, 6 and 7. The GCSE language performance for each group increased between 2003 and 2006.
7. These analyses provide support for GCSE grading standards having been maintained over time. Neither of these analyses supports the contention that it has become more difficult to achieve a GCSE language grade C over recent years. However, that does not address the claim that GCSE languages are and have been more demanding than most other subjects.

## **Attainment-based comparability**

8. The QCA view of comparability, which has formed the basis of the studies we have published over the last 10 years, is that the standard of performance of a particular award lies in its intellectual demand – the level of knowledge, skill and understanding acquired by students – more specifically, the quality of candidates' attainments at each key grade boundary. This is an 'attainment-based' conception of comparability: for examination standards to be considered comparable, the attainments of students at equivalent grades must be similar.
9. The 'attainment-based' conception of comparability is implicit in the method used in GCSE and A level awarding meetings at which decisions are made about the marks needed to achieve particular grades. The code of practice governing these meetings makes it clear that 'each boundary must be set using professional judgement. The judgement must reflect the quality of candidates' work, informed by relevant technical and statistical evidence.' Fundamentally, awarding meetings are required to ensure that the quality of work describing a particular grade this year is the same as that of the previous year.
10. The 'attainment-based' conception of comparability is less straightforward when comparing standards across subjects (as opposed to comparing standards from one year to the next within the same subject), since it requires us to think in terms of general kinds of knowledge, skills and understanding. Pitching standards at the right

level, to ensure comparability between subjects, happens primarily when subject criteria are developed – rather than during awarding meetings – through the selection of appropriate content and through the articulation of appropriately phrased grade descriptions.

## **Aptitude–based comparability**

11. There is, though, a different approach to comparability. The ‘aptitude-based’ conception of comparability holds that two examinations may be seen as comparable if students of a certain ‘calibre’ have an equal chance of achieving a particular grade in any examination (either within subject areas or between them). One examination is harder than another if the results of a (representative) group of students taking both examinations are worse in it. This is the approach to comparability on which the ALL proposal is based.
  
12. To test its practical application we asked AQA, Edexcel and OCR to calculate what changes there would be to their 2007 GCSE French results if the grades were based only on the relationship between candidates’ key stage 3 test scores and their results in GCSE mathematics. The analyses indicated that the changes would be marked at the higher grades. About half the candidates presently awarded a grade B would gain a grade A as the threshold mark or performance standard for a grade A would have to move down by about half a grade width. There would be a similar effect at grade C.
  
13. We also asked the awarding bodies to carry out a similar study using GCSE art instead of GCSE French to see what effect there would be on aligning a very different subject with mathematics. Here the impact at the higher grades is in the opposite direction. About a third of the candidates presently awarded a grade A in art would gain a grade B as the threshold mark for a grade A would have to move up. There would be a similar effect at grade C with perhaps a third of those candidates being reclassified as Ds.
  
14. These estimates of predicted changes are summarised in the table below, using the summer 2007 provisional GCSE results for the UK as a baseline.

	%A*–A	%A*–C
French summer 2007	23	66
French aligned with mathematics	30	77
Art summer 2007	22	73
Art aligned with mathematics	18	61

15. As can be seen, the estimated proportions achieving grades A\*–A in summer 2007 in the two subjects were very similar. Adjusting the results statistically to bring them into line with mathematics creates a 12 per cent difference. The summer 2007 results show 7 per cent more candidates awarded grades A\*–C in art than in French. After applying the adjustment there are an estimated 14 per cent more candidates awarded grades A\*–C in French.

## Discussion

16. In this country, performance standards are based on professional judgements made by those within the subject community. All those who help set the criteria on which the syllabuses are based and set the grade boundary marks are drawn from that community.
17. Most of the public would accept that being examined in French and in mathematics are not comparable experiences. To lower our expectations of what performance to expect for a particular grade in French on the basis of a statistical comparison with mathematics is likely to produce accusations of diluting intellectual demands and reduce confidence in the results.
18. If we want to base our examination system on a statistically-derived ‘aptitude-based’ conception of comparability then we would have to select a year and align all subjects on that basis from that time. In this system, the proportion of C grades awarded each summer would be pre-determined no matter how poor the real examination performances of the students were. Its introduction would lead to some radical implications for grades awarded in the apparently ‘easiest’ and ‘hardest’ subjects: the ‘easiest’ (such as English and art) would see a substantial fall in the proportion of higher grade candidates, while the ‘hardest’ (such as French and German) would see a substantial rise.

19. Perversely, 'aptitude-based' analyses put Urdu and some other languages into a different category from French. An 'aptitude-based' alignment with mathematics would raise the proportion of higher grade candidates in French but lower the proportion in Urdu. This illustrates neatly a major problem with introducing the 'aptitude-based' approach: two languages where grade C candidates presently produce work showing comparable knowledge, skill and understanding would be realigned so that better work would be needed in Urdu than in French for the same grade. In this example the realignment would be overt because the two examinations are relatively straightforward to compare. However, the same change in grading standards would happen in subjects where the comparison of knowledge, skill and understanding is more difficult to make, such as English and ICT. Attempts to introduce exceptions for subjects such as Urdu would inevitably undermine the basis of the 'aptitude-based' approach.
20. Few countries in the world use such an 'aptitude-based' system. The Australian states and territories seek to achieve statistical comparability across subjects, at least for university entrance purposes. The independent committee set up by QCA to review techniques for monitoring comparability reported that 'they achieve this by making the rather heroic assumption that all examinations are measuring essentially a common dimension.' We might think of the common dimension as 'general aptitude'. Thus in Australia, if the marks of all students taking art, geography and mathematics show that most students gain their highest marks in geography and their lowest marks in mathematics, it is concluded that geography is the easiest subject and that mathematics is the hardest of the three. The three distributions are then adjusted, so that mathematics has the lowest pass mark and geography the highest. Both the unadjusted and adjusted results are presented to students; the adjusted results presented as a ranking are used in particular for selection to university, while the unadjusted results for each student are used to report achievements in each subject.
21. In practice it is not clear that the scaling process used persuades parents and students that there are no problems regarding comparability between subjects. The assumptions underlying this adjustment are contentious and have not been accepted in England. Adopting the process would make it very difficult for users to compare students who were awarded grades in these subjects before alignment with those examined after the alignment. Users would also struggle to understand what knowledge and skills to expect from students awarded particular grades, as the

awards would not be based simply on judgements of the quality of work in the subject but on the general aptitude of the students as well.

## Conclusion

22. The present examination system is not based on an aptitude-based conception of comparability and its adoption would create a major threat to public confidence in students' results. We have similar concerns about applying a new performance standard to languages but not to any other subjects, some of which might also want to claim special status. We do not have evidence that there have been significant changes to grade standards in GCSE languages in recent years. In conclusion, we do not therefore, recommend any adjustment to national grade standards in GCSE languages.
  
23. Instead we should focus on improving levels of teaching and learning in modern languages in order to gain students' commitment and raise performance. In relation to GCSE performance the Languages Review recommended reconsideration of the methods of assessment for speaking and listening, given the restricted nature of the current arrangements. Following an investigation QCA has introduced changes to the GCSE criteria. These changes will enable students to show their knowledge and skills in different ways. QCA is undertaking further work to investigate and illustrate the new forms of assessment. The revised GCSE criteria also allow for more choice, so that teachers will have opportunities to tailor the content of courses to be more relevant to students' interests.
  
24. We will also keep working with the awarding bodies to ensure that grade standards in GCSE languages over time continue to be properly maintained in the future and that any small deviations from the national standard are detected and then corrected.