Comparability of Different GCSE and A Level Subjects in England: An Introduction

ISC Working Paper 1

December 2015

Ofqual/15/5797

Contents

Suggested citation:

Ofqual (2015a) *Comparability of Different GCSE and A Level Subjects in England: An Introduction: ISC Working Paper 1*. Coventry, the Office of Qualifications and Examinations Regulation.

This report was written by Dennis Opposs (Standards Chair).

Introduction

The comparability of exam standards can be viewed in different ways. Concerns are sometimes raised about whether standards of GCSEs or A levels are being maintained over time. On some occasions, there is disquiet about the comparability of standards between the exams offered by different exam boards within a subject. For some years now there have been issues raised about the comparability of GCSEs and A levels with vocational qualifications that full-time students can take instead. In recent times, it has been suggested that our GCSEs and A levels do not stand comparison with the equivalent exams used in some educationally high-performing countries.

There is a fifth dimension to comparability. 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. We are investigating technical, practical and policy issues in relation to the comparability of different GCSE and A level subjects in England, and we intend to establish our position once we have heard the views of others.

To help throw light on the present position, stimulate informed debate and help us decide what to do, we have produced a set of six working papers. This working paper considers why the comparability of different GCSE and A level subjects is an important issue to consider, focusing particularly on possible effects on the balance of the curriculum taken by GCSE and A level students. It also draws on the other working papers to provide a brief introduction to relevant considerations.

Working Paper 2 (Ofqual, 2015b) is a literature review of research conducted in this area that describes different conceptions of inter-subject comparability, relevant methods for investigating it, the outcomes of some of the studies that have been published and the views of many who have worked in this area.

Working Paper 3 (Ofqual, 2015c) is based on a statistical analysis that looks at the consequences of aligning standards based on ‘difficulty’ estimates from a statistical technique known as Rasch. It considers what the impact would be in different subjects of aligning standards in terms of grade boundary marks, proportions of grades and performance standards.

Working Paper 4 (Ofqual, 2015d) provides a broad overview of whether other countries around the world address the comparability of subjects in their assessments and, if so, what they do. The methods that some countries use are identified along with public perceptions of their use.

Working Paper 5 (Ofqual, 2015e) describes the positions that our predecessors have taken over the last 20 years in relation to the comparability of subjects. Although, from time to time, the exams regulator of the moment acknowledged that something might be done, things have always been left as they were.

Working Paper 6 (Ofqual, 2015f) identifies policy options that we could adopt in relation to the comparability of subjects and the benefits and drawbacks associated with each. Broadly, the possible options are: exam boards adjusting the grades they issue; leaving certificated grades alone but adjusting the way these grades translate into measures such as the UCAS tariff and the Department for Education (DfE) school performance measures; leave things as they are.

Why does inter-subject comparability matter?

It can be argued that the exams system in England is largely fulfilling its key purposes of:

- certificating A level students in a way that permits higher education institutions (HEIs) such as universities to decide which students they want to admit to their courses on the basis of information about subject attainment and, similarly, allows employers to decide which students they want to employ; and
- certificating GCSE students so that information on their subject attainment allows progression routes to operate sensibly, and generates data that the DfE can use to measure the performance of schools.

Are there, though, any indicators that might predict a risk to the exams system and to which we should have regard? Four risk indicators are suggested below.

- Subject difficulty – real or perceived – might dissuade students from choosing some subjects or might affect decisions that schools make on behalf of their students. Such concerns have been raised by linguists, and science, technology, engineering and mathematics (STEM) subject specialists and headteachers.
- HEIs might select the wrong students for their courses because they assume that A level grades for different subjects can all be counted as equal.
- Using exam results from different subjects for teacher or school evaluation might lead to poor decisions about, for example, which teachers to promote or which school a parent decides is best for his or her child.
- Incomparability between subjects might have a deleterious effect on public confidence in the exams system.

The first two risk indicators are explored in a little more detail below.

Subject difficulty affecting decisions made by schools

In 2014, we carried out an anonymous online survey, primarily aimed at schoolteachers. The self-selecting respondents were asked whether they had experienced, first-hand, a range of listed activities within the last year. Over 50 (less than 10 per cent of the total responses) suggested that teachers and schools made subject and qualification choices for individual students based on what was in the best interests of the school rather than the student. Responses included comments that students were 'required' to take specific subjects which were perceived to be easier, with examples including GCSEs in religious studies and physical education. The reverse also applies, with instances of high-attaining students being encouraged to take more traditionally 'academic' subjects when they might have wished to experience a greater breadth of subjects.

Relevant here are statistics[1] showing GCSE outcomes by subject, based on students' prior attainment in the Key Stage 2 tests taken 5 years earlier. Taking the pupils who achieved Level 4b[2] in Key Stage 2, the DfE data show that the proportion achieving at least a grade C in a variety of subjects is:

GCSE French	44%
GCSE German	47%
GCSE physical education	58%
GCSE religious studies	60%

Faced with such data, headteachers may advise many students with such a level of prior attainment not to choose a language if instead they can opt for a subject that, according to national-level statistics, provides a better chance of achieving a grade C. Of course, there are other performance measures that influence headteachers' actions – see below for the English Baccalaureate (EBacc) – but until now the measure of 5+ grades A* to C (including English and mathematics) has been very powerful. The new Progress 8 measure[3] may also incentivise headteachers to find those subjects that appear to produce the best improvement in results between Key Stage 2 tests at age 11 and GCSEs at age 16.

[1] www.raiseonline.org/documentlibrary/ViewDocumentLibrary.aspx .
[2] Level 4b indicates an average of mid-Level 4 achieved across the Key Stage 2 tests in English, mathematics and science.
[3] The Progress 8 measure for schools will be based on students' progress measured across eight subjects: English; mathematics; three other EBacc subjects (sciences, computer science, geography, history and languages); and three further subjects, which can be from the range of EBacc subjects, or can be any other approved arts, academic or vocational qualification.

Although it may well be that take-up of GCSE modern, and perhaps ancient, languages has been affected by such interpretations of data, it is not clear that such perceptions apply to other GCSE subjects. They may, though, have played a part in the growth of vocational qualifications where the Wolf report (Wolf, 2011) says that, "young people [are] encouraged to take [vocational] qualifications at age 14-16 […] for reasons which have nothing to do with the pupils' own best interests." (page 44) This set of working papers, though, restricts itself to comparability within GCSEs and A levels, not between GCSEs and A levels and other qualifications.

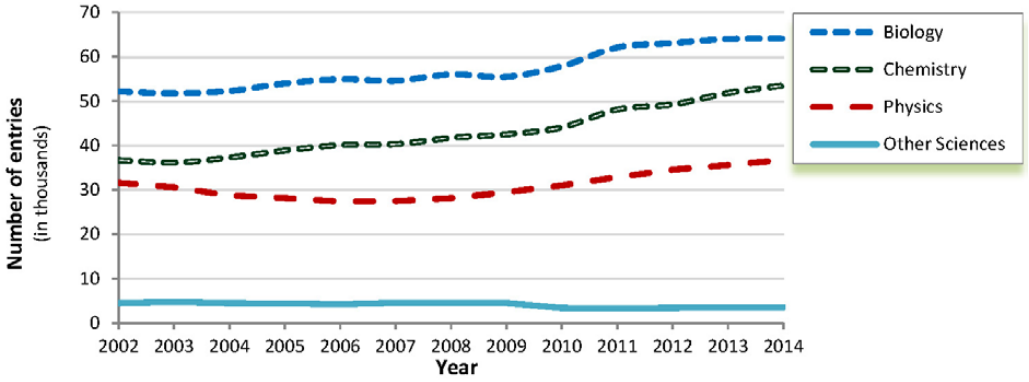
Of course, there are interpretations of the above data other than languages are harder than physical education. For example, these data might arise if the students who had achieved Level 4b in Key Stage 2 who took physical education were typically more interested in the subject and more motivated to succeed than the students who took French. The generalisation is also flawed; the same dataset shows that the highest performers (Level 5a) have a better chance of achieving a grade A or A* in French than in physical education. However, it is the assumption that languages are harder than physical education that has led to a particular behaviour in some schools in England, affecting the Key Stage 4 curriculum of many students.

Higher education institutions selecting students

The position is different at A level in that typical students are more likely to know their strengths and weaknesses and have future progression routes in mind than their GCSE counterparts. Concerns have been raised in recent years by the teaching community in French, German and Spanish that these subjects are more severely graded than others. The analysis in Working Paper 3 shows these subjects to be hard (in a statistical sense), although not generally as hard as the sciences.

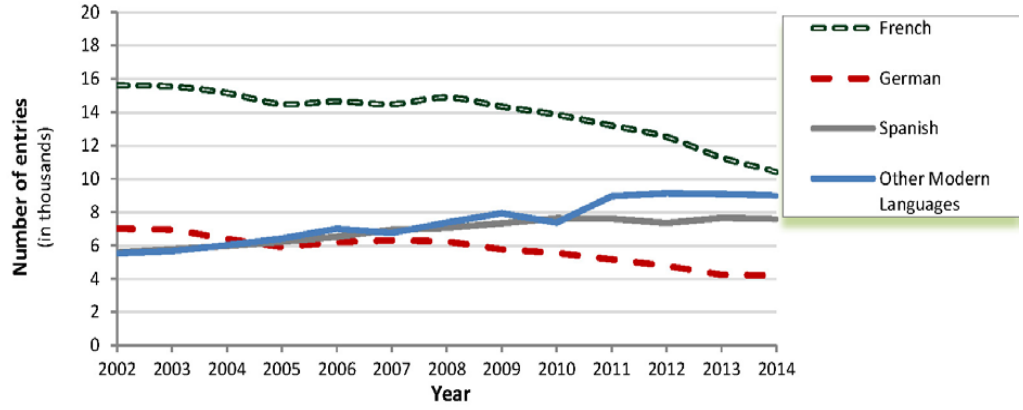
Organisations such as the Institute of Physics have been concerned in recent years about the take-up of A levels in the sciences and mathematics, fearing that their perceived severity is affecting student choice. Although no action has been taken to adjust grading standards, entries for A levels in these subjects have risen noticeably in the last few years.

A level sciences entries, 2002 – 2014
(Source: Joint Council for Qualifications)



The position in languages is different with entries for French and German in decline.

A level modern languages entries, 2002 – 2014
(Source: Joint Council for Qualifications)



These data seem to indicate that changes in entry for a subject over time can change even when no change has been made to its grade standards.

Of course, students aiming to study a language or a STEM subject at university have to choose at least one relevant subject at A level, even if they believe they might get better grades in others. That is because offers by universities to potential students normally specify more than just UCAS points, for example 300 points including at least 180 in mathematics or two As and a B with at least an A in French. So, typically, competition for places on, say, a chemistry degree is among students who have studied at least some similar subjects.

There are, though, degree courses such as law and social psychology where the students applying will have less in common in what they have studied at A level. That may well result in a grade B in one subject being treated as worth a grade B in a very different one by admissions tutors. Where the same grades in different

subjects are being treated as equal currency then, if it is believed that some subjects are more severely graded, there might be more of a justification for action.

What could be done about inter-subject comparability?

Working Paper 5 describes the different positions that our predecessors adopted in relation to the comparability of different GCSE and A level subjects. The aim of this programme of work is to allow us, on the basis of as much information as possible, and considering fully the implications for the exams system of any possible changes, to determine our position.

Among the policy positions that we could take if we decided that positive action was called for, there are two broad positions:

- exam boards adjust the grades they issue; or
- certificated grades are left alone but we advise that there is an adjustment made to the way these grades translate into measures such as the UCAS tariff and the DfE school performance measures.

Working Paper 6 describes in considerable detail a set of policy options with respect to the comparability of different subjects, and it identifies issues related to if we adopted any of them.

Changing grade standards

Working Paper 3 provides data, for a selection of GCSEs and A levels, for what the practical impact would be on percentage outcomes in grades of aligning standards using a statistical approach consistent with a policy of exam boards adjusting the grades that they issue. For example, in A level physics the proportion of students awarded grades A* and A would rise from 31 to 48 per cent.

So far, the exam boards have not made any attempt to adjust certificated grades in England to take account of inter-subject comparability other than occasionally within very similar (cognate) subject areas, for example French and German. Nor has UCAS or the DfE applied different points conversions to the same grades in different subjects. There is a question here about which organisation has ultimate responsibility for such actions. There is presently no requirement in our regulations for the exam boards to align subjects. Presumably, if the owner of a secondary measure wanted to use different conversion rates for different subjects in its measure it would be free to do so.

Changing the exchange rate for grades

Grades are 'converted' by others in the system. DfE uses them for performance tables to hold schools to account, and universities use them to select students for higher education courses. Generally, the 'exchange rate' for these conversions does not vary by subject. But we could vary the exchange rate according to the difficulty of the subject. And different users could apply their own exchange rates, depending on how they were using the grades.

For example, DfE could change the way that it converts GCSE grades into performance points so that, unlike now, the exchange rate varied from subject to subject. A similar opportunity exists for UCAS and its tariff. Deciding exactly what changes to make, though, would be hugely challenging. As Working Paper 2 says:

> […] although thinking has advanced considerably over time, we still see huge disagreements concerning how best to define and conceptualise inter-subject comparability, let alone how best to monitor it, let alone how best to respond to monitoring outcomes (p. 40).

That does not mean, of course, that the owner of a standard or dataset – be it us or DfE or UCAS – could not decide that, for its purposes, one particular definition would apply. Even having done that, there would be difficult decisions to make about some of the details – whether the adjustments make easy subjects appear harder or vice versa, and what to do about apparent anomalies in the statistics for subjects like Urdu (see Working Paper 6, p. 24).

This paper started by giving the different dimensions of comparability – over time, between exam boards in a particular subject, across qualification types, internationally and across subjects. Decisions to adjust certificated grades in some subjects might be seen as a failure to maintain the comparability of grade standards over time in those subjects. Maintaining standards over time is one of the objectives that Parliament has given us (see the Education Act 2011).[4] Presently, we fulfil that objective with regard to GCSEs and A levels using a comparable outcomes approach to the carrying forward of grade standards. It might be argued that this approach could be compatible with some inter-subject comparability adjustments which could be made.

[4] "The qualifications standards objective is to secure that - (a) regulated qualifications give a reliable indication of knowledge, skills and understanding, and (b) regulated qualifications indicate - (i) a consistent level of attainment (including over time) between comparable regulated qualifications, and (ii) a consistent level of attainment (but not over time) between regulated qualifications and comparable qualifications (including those awarded outside the United Kingdom) which are not qualifications to which this Part applies."

Changing behaviour through influence

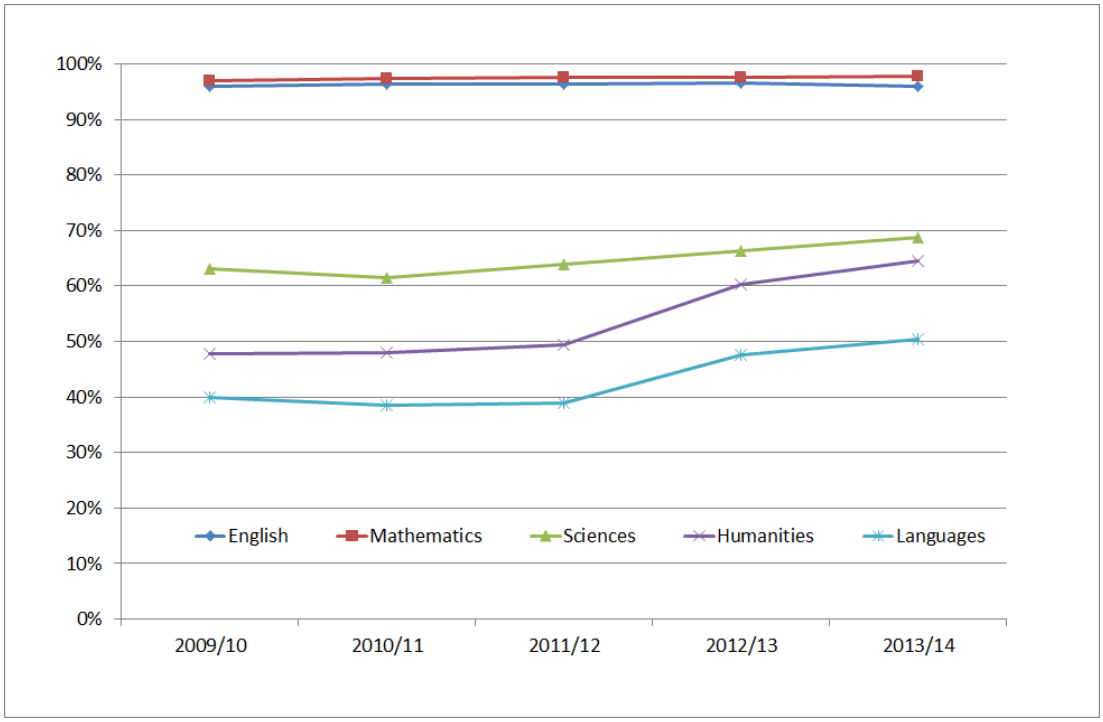
Changes in recent years within Key Stage 4 in England have included increasing numbers of students taking non-GCSEs and decreasing numbers taking GCSEs in French and German. Government actions have affected these trends, but not by changing the comparability of standards of GCSE subjects.

Government actions have included reducing the value in school performance measures of non-GCSEs and introducing the EBacc, a performance measure for schools. Students achieve the EBacc if they gain GCSEs at grades A* to C in English, mathematics, history or geography, two sciences and a language.

The EBacc was announced by the Government in autumn 2010. Summer 2013 was the first time that students' choice of subjects was fully reflected in the GCSE results. In summer 2013, the decline in entries for languages was reversed, the total of language entries being a five-year high. Compared with 2012, French entries were up 16 per cent, German entries up 9 per cent and Spanish entries up 26 per cent. Also in 2013, the number of entries to history was at its highest for at least 16 years, whilst the number of geography entries was at its highest for nine years. Although entries for all these subjects remained broadly steady in 2014, the continuing impact of the introduction of the EBacc is shown in the graph below.

Percentage of students entered for components of the EBacc

England, 2009/10 – 2013/14, state-funded schools only (Source: DfE)



These changes in entries happened without any changes being made to grade standards or to the rate at which those grades were converted into DfE measures.

The Government has recently set out its goal for the vast majority of Key Stage 4 students to enter the EBacc. Such a change would again affect the take-up of several subjects.

International experience

As Working Paper 4 makes clear, outside England there is a minority of jurisdictions where statistical corrections are made to adjust for subject severity.

For example, in New South Wales, student achievement in Stage 6 (Years 11 and 12) is reported in two ways: the Higher School Certificate and the Australian Tertiary Admission Rank. Scaling is used based on the 'equal achievement' principle - when the same group of students takes a pair of subjects, then the average performance of the group on each subject should be roughly the same. Scaling adjusts the raw scores in all subjects so that the scaled scores used in the Australian Tertiary Admission Rank are comparable.

The Hong Kong Examinations and Assessment Authority equates grade standards across all subjects at each level in its Diploma so that grades across subjects are comparable. It uses a group ability index when grading elective subjects. The group ability index is calculated for an elective subject based on those students' results in the four core subjects and the correlations between the subjects.

In the Taiwan Basic Competency Test, taken at the end of junior school, statistical models (using item response theory) are used to convert the pupils' raw scores to a scaled score in each of five subjects to enable direct comparisons between the scores obtained in each subject.

Of course, many jurisdictions have a far more prescriptive and broad post-16 curriculum than we have in England, and so there may be less of a need to consider the relative demands of different subjects.

As Working Paper 4 points out (pp. 14–17), international experiences show that the complex methods used to scale exam results can be impenetrable to the public, leading to concerns about fairness and transparency. Scaling in this way doesn't always remove unintended student behaviour. In Australia and Cyprus there have been reports of students trying to 'game' the system by opting for subjects that are usually scaled up and avoiding those which are scaled down.

We must also remember that any action taken in England has to be implemented in the same way by each of the exam boards that provide the qualifications, a practical hurdle that other jurisdictions do not have to face.

References

Education Act 2011. (c21), London: HMSO

Ofqual (2015b) *Inter-Subject Comparability: A Review of the Technical Literature: ISC Working Paper 2*. Coventry, the Office of Qualifications and Examinations Regulation.

Ofqual (2015c) *Inter-Subject Comparability of Exam Standards in GCSE and A Level: ISC Working Paper 3*. Coventry, the Office of Qualifications and Examinations Regulation.

Ofqual (2015d) *Inter-Subject Comparability: an International Review: ISC Working Paper 4*. Coventry, the Office of Qualifications and Examinations Regulation.

Ofqual (2015e) *A Recent History of Regulatory Perspectives on Inter-Subject Comparability in England: ISC Working Paper 5*. Coventry, the Office of Qualifications and Examinations Regulation.

Ofqual (2015f) *Exploring Implications of Policy Options Concerning Inter-Subject Comparability: ISC Working Paper 6.* Coventry, the Office of Qualifications and Examinations Regulation.

Wolf, A. (2011). *Review of vocational education: the Wolf report*. London. Stationery Office

We wish to make our publications widely accessible. Please contact us at publications@ofqual.gov.uk if you have any specific accessibility requirements.

This publication is available at www.gov.uk/ofqual.

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation

Spring Place
Coventry Business Park
Herald Avenue
Coventry CV5 6UB

2nd Floor
Glendinning House
6 Murray Street
Belfast BT1 6DN

Telephone 0300 303 3344
Textphone 0300 303 3345
Helpline 0300 303 3346