Research Report DFE-RR203

The evaluation of the impact of changes to A levels and GCSEs

Final report

AlphaPlus Consultancy Ltd

Department for Education

Executive summary

Introduction

This report presents findings from the independent evaluation of the impact of changes to A levels and GCSEs in England. At A level, these changes were introduced to all subjects except mathematics in September 2008. The changes included a move from six units to four for the majority of subjects (physics is an exception)G, the introduction of greater stretch and challenge at A2, and the introduction of the A* grade. From September 2009, changes were introduced to GCSEs in all subjects except English, mathematics, ICT (information, communication and technology) and the sciences. Changes to GCSE English, mathematics and ICT were introduced in September 2010. At GCSE, controlled assessment[1] and unitised assessment were common aspects of the new specifications.

The report follows three rounds of fieldwork with case-study centres and wider stakeholder groups designed to ascertain their understanding and perceptions of the changes, and their attitudes towards them. It builds on findings from the statistical data reported on in spring 2011,[2] and reports on early indications of the impact of the changes and the issues that have arisen.

The initial round of data collection during 2010 concentrated on six GCSE subjects (French, geography, health and social care, history, media studies and Spanish – for which new specifications were introduced in September 2009) and on six A level subjects (English literature, French, geography, media studies, physics and psychology). A change of focus[3] following the Department for Education's (DfE) take-over of the contract for the evaluation has meant that, at GCSE, data was collected for English/English language and mathematics (for which new specifications were introduced from September 2010) instead of for health and social care and media studies. At A level, data was collected for history instead of media studies.

---

[1] The purpose of controlled assessment is to assess those aspects of a subject that cannot be easily assessed by external examination. These aspects include research, planning, investigation, analysis, collaborative working, and presenting ideas and arguments supported by evidence. Controlled assessment is designed to encourage a more integrated approach to teaching, learning and assessment, and to enable teachers to confirm that students carry out the work involved.

[2] DfE (2011) *The evaluation of the impact of changes to A levels and GCSEs – second interim report*, reference DFE-RR170 www.education.gov.uk/publications/RSG/NewRsgPublications/Page1/DFE-RR170

[3] English and mathematics, the new GCSEs added to the evaluation, are included in the national curriculum for Key Stage 4 (2007). The change of focus at GCSE and A level was made at the request of the DfE, to reflect the coalition government's interest in the 'core' subjects.

Aims of the evaluation

The impact of the changes was considered under three themes:

- stakeholders' perceptions of the new GCSEs and A levels
- the impact of the changes on teaching and learning
- the impact of the changes on centre behaviour and management of change (including assessment).

This report focuses on building an understanding of the impact of the changes to specifications for A levels and GCSEs across the different phases of implementation. The specifications have been implemented at incremental points within a two-year timeframe, and the interpretation of the data may therefore depend on where stakeholders are located on the change continuum for each qualification and/or subject.

Methodology

Data sources

The following data sources have been used for the evaluation across the three reporting phases:

- Centre research study (CReSt) data (provided by QCDA from the CReSt project)
- Literature review: the Qualifications and Curriculum Development Agency's (QCDA's) research evidence management system (REMS) database (March 2010)
- Awarding body data archive (ABDA) – awarding organisation data supplied by the Office of Qualifications and Examinations Regulation (Ofqual)
- Official statistics – statistical first releases (SFR) – from the DfE Research and Statistics Gateway
- Joint Council for Qualifications (JCQ) data
- National pupil database (NPD) – including census, pupil level annual school census (PLASC) and examination results
- data from 17 case-study centres – including interviews or focus groups with senior leadership teams, heads of department, subject teachers, examination officers, and students
- interviews with 19 wider stakeholder representative groups[4] and 35 responses to an online survey from 29 English higher education institutions (HEIs)[5]

[4] Employers, subject and professional associations, FE representative organisations, trade unions, workforce development agencies and awarding organisations.

- awarding organisation and Ofqual document review (updated in 2011 to include new subjects)[6]
- centre online survey data – 44 centres responded in total (6.4% response rate): 19 completed responses to the senior leader questionnaire, 97 responses from 41 different centres to the subject-specific questionnaires.

Summary of findings

Across the three phases of evaluation fieldwork undertaken there have been varied and sometimes contradictory views from schools and other stakeholders on the outcomes and impact of the changes to the specifications.

The data from awarding organisations and national statistics provide information about trends in participation and grade outcomes. Data from the case-study centres and stakeholder surveys and interviews capture individual experiences, viewpoints and assessments of the impact of the changes, and reflect differences in ethos, context and what are felt to be the main drivers, and therefore priorities, for a particular education phase.

Stakeholders' perceptions of the impact of changes to A levels on teaching and learning

Stretch and challenge

The introduction of 'stretch and challenge' in A levels and the reduction from six to four units for most subjects[7] have been welcomed by the majority of stakeholders. Teachers felt that A2 English literature, geography, psychology, history and French were more challenging, and they considered them better preparation for higher education –for example, in the way they encouraged independent learning. There were some exceptions in terms of specific skills development for a subject – for example, some geography teachers felt that there was not enough emphasis on the research skills required for university study. The views were more mixed for physics, depending in part on where students were progressing to and how the subject related to their choice of further study. In those subjects that were affected, the

[5] An invitation to take part in an online survey was sent to 103 HEIs in England and followed up with reminder emails and telephone calls; Supporting Professionalism in Admissions (SPA) also promoted the survey at one of their events and in a newsletter. Response rate of 27% from the original 103 HEIs invited – one additional HEI also responded.

[6] The review of the awarding organisation documentation follows the logical development process of the revised qualification: the original qualification criteria and the resulting awarding organisation specification(s); the revised subject criteria for each qualification and the responding awarding organisation specification(s); the sample assessment materials, along with relevant mark schemes, and any other qualification-specific awarding organisation guidance material aimed at teachers and learners.

[7] With the exception of physics, where there continue to be six units in the new specifications.

decrease from six to four units at A level was seen as positive in terms of ensuring greater depth and breadth of study.

Prior to the examination results in summer 2010 for the new-specification A levels, centres had been concerned that there would be a negative impact on students' grades because of what they saw as the greater challenge of some of the new specifications, and because of the uncertainty about how this would be assessed in practice. The cause of this concern, in part, was that centres were not always aware of the role that Ofqual and the awarding organisations had played in ensuring that the first students to take examinations under the new specifications were not disadvantaged. Lack of familiarity with the new specifications and assessments meant that teachers felt less able to predict how the examination questions would assess and reward performance.

A clear distinction was not always made between opportunities for stretch and challenge within teaching and learning – i.e. the higher-order skills developed within the course itself – and how these skills are to be assessed and evidenced in responses to examination questions. Stretch and challenge has also been interpreted in a variety of ways by centres, and there continues to be a degree of uncertainty and ambiguity about what is involved. The stakeholders interviewed had mixed views on the extent to which the new specifications had made a difference – and, if they had made a difference, on the degree to which this had been recognised and the need for change understood. Drawing on findings across the three rounds of fieldwork, however, the majority of centres felt that there was greater stretch and challenge in the new specifications at A2, but that this was not always equally reflected in the assessments, possibly as a result of the different ways in which awarding organisations can introduce stretch and challenge. There was also disagreement between and within subjects as to how best to deliver stretch and challenge and, indeed, the extent to which stretch and challenge was possible for all students.

The structural aspects of the changes to the majority of A levels (i.e. all except physics), the organisation of content and the mode of assessment (for example, course work or external assessment) were not necessarily perceived as encouraging greater stretch and challenge. What was important was the effect structural changes had on the teaching and assessment of higher-level skills. Many geography teachers, for example, considered that the removal of the coursework element had reduced the level of stretch and challenge because it had been through coursework that students developed the fieldwork skills required for university study. Centres often discussed stretch and challenge in the context of A levels in terms of developing independent learners. Here the mode of teaching and learning was an important factor, but a high level of teacher input and initial support was needed if the student was to develop the underpinning skills (such as research skills) required.

The specifications criteria require that A2 assessments must include appropriate demand by:

- the use of a variety of stems in questions
- ensuring connectivity between sections of questions
- requiring extended writing in all subjects (except where it had been agreed with the regulatory authorities that this was inappropriate)
- using a wide range of question types to address different skills
- including synoptic assessment.

Awarding organisations have, accordingly, introduced stretch and challenge in one or more of a number of ways:

- in the qualification syllabus (for example, by increased content or by placing more emphasis on higher-order skills/concepts)
- by making changes to the assessment objectives and/or relative weightings
- in the paper setting (in the form of amendments to the nature of the questions/tasks)
- in the marking criteria
- at the grade-awarding process.

The different approaches depend in part on differences in the nature of subjects and the way skills and knowledge are organised. For example, in modern foreign languages knowledge, understanding and skills are closely linked, and synoptic assessment promotes stretch and challenge.

Differences of opinion across subjects on whether the 'application' of subject skills and knowledge created greater stretch and challenge were largely expressed in terms of the methods of enquiry for a discipline. For example, for physics the greater emphasis on 'application' was thought by some teachers to have lessened the more important basic principles of the discipline that come with learning and understanding equations and more-complex mathematical problems. For other centres 'application' had proved too challenging, and they had changed to a different specification that was considered more accessible for their students.

Teachers felt that they needed time to amend their teaching strategies and that students needed time to adjust to new approaches in their learning in order to develop higher-level and independent learning skills. Independent and synoptic learning and understanding are often considered more 'difficult' because they may introduce new ways of thinking and ordering knowledge. How much additional time students require to master these 'higher-

level' ways of thinking about their subject clearly depends on their starting point and what they are used to. Additionally, independent learning approaches need, initially at least, to be scaffolded and supported, which is labour intensive for teaching staff. This was particularly evident in the extended project qualification, which required considerable staff resourcing, but which was also widely reported to be of benefit to student and staff development, with wider stakeholders, including HEIs, recognising the benefits of the qualification. The A* grade at A level is perceived by both teachers and students as recognising high-attaining students. Some universities now require an A* pass for certain courses.

Coursework

Although teachers considered coursework a burden, in terms of their workload, they also felt it offered assessment (with a formative and summative element) that supported development. Teachers, managers and students suggested that A level coursework performance was a better indicator of a student's ability than written examinations. Physics apart, the majority of A level students' subject- and non-subject-specific comments suggested that they would wish to retain coursework as part of the assessment regime.

Progression

Progression from AS to A2 was considered by many case-study centres to be a larger jump than in the previous specifications. During the 2011 follow-up visits for English literature, history and geography, for example, those who responded thought that the AS was preparing students less well for A2 than before. Analysis of examination data indicated that, for most of the subjects looked at in this study, proportionally fewer students continued to A2 for 2010 completion (new specification) than had continued to A2 for 2009 completion (legacy specification).

Student engagement

The majority of the heads of Key Stage 5 (KS5) reported in autumn 2011 that there was generally greater engagement in the new A levels. Overall, they considered this was less to do with specific content and more to do either with the development of skills or with the structure of the course.

Grade outcomes at A level

Two important caveats need to be placed on findings from the statistical data:

- summer 2011 data is for entries rather than for candidates – this may mean that the effects of re-taking (the entire qualification) cloud information about overall outcomes

- no data is available yet for candidate characteristics, so it is impossible to separate out whether any changes in attainment are due to changes in the assessment or to changes in the candidature, or are affected by both.[8]

In general, entries at A level appear to follow the existing trends and not to have been greatly affected as yet by the introduction of the new-specification awards. For all new-specification A level subjects in 2011, entries have decreased very slightly since 2010 (down from 703,000 to 696,000), following four previous years of increasing numbers since 2007.[9] However, there are variations from subject to subject, with entries for physics, for example, increasing.

Between 1996 and 2009, A level attainment in most subjects broadly followed the same improving trend (an average increase of 0.06 of a grade per year, with short-term declines from the trend at times of specification change), and since 2003 entries have also been increasing in most subjects.

In 2010, when the first awards of the latest specification change were made, attainment saw a plateauing of grades following the long period of increase (a decline from the prior trend similar to, but smaller than, when previous changes to curricula had been made). When all new specification subjects were combined, the same proportion of students obtained the A grade in 2010 (including those awarded the new A* grade) as had received an A grade in 2009 (24.6%), while from 2003 to 2009 an average of 0.7% more students received an A grade year on year.

Considering the new specification A levels as a whole, the plateau effect seen in the A level results in 2010 appears not to have been repeated in 2011. From 2010 to 2011, the upward trend in average grades that had been seen up to 2009 was largely restored. Whether this is because the 2011 cohort had higher prior attainment at GCSE than their 2010 counterparts cannot be determined until NPD data becomes available (from March 2012). For the four individual subjects considered, however, there are variations from this overall trend, as might be expected.

The plateauing of grades at A level seen in the 2010 results did not fully reflect the perceived increased level of challenge that teachers had reported before the results from the summer 2010 examinations became available. Many teachers had expected to see a drop in grade

[8] Data about candidates would enable the effects relating to the candidature's demographic and prior attainment characteristics to be separated from effects relating to the assessment, and allow investigation of whether candidates of different backgrounds or prior attainment have been affected differently by the new specification.

[9] Since students of school age (i.e. 16–18) are of most interest in this report, the numbers are based on the SFR data.

outcomes after the change of specification, especially at A2. In the event, the plateau in grades is likely to be due to a combination of factors: the challenges of the new specification and the impact of measures introduced by Ofqual for the 2010 A level series (placing an effective limit on the extent to which grades could rise or fall) to ensure that the first candidates to take the new-specification A levels examinations were not disadvantaged. This presents awarding organisations with a dilemma: strong guidance from Ofqual to use predictive matrices for annual GCSE and A level awards based on prior attainment data increases the emphasis at grading on attainment data from the previous year's cohort and so may over time compromise the criterion-related element which currently forms part of grading decisions at A level and GCSE. Additionally, such a move towards grade outcomes comparable with those in previous years would also be likely to prevent GCSE and A level grades increasing in future as a result of improvements in teaching and learning, as such improvements could be interpreted at grading as 'grade drift' (i.e. teachers becoming familiar with the new requirements rather than improvements in teaching and learning *per se*).

Throughout the two years of the new-specification A levels, the difference in levels of performance between females and males has remained largely unchanged: females consistently achieve higher grades.

A number of centres commented on the rise in numbers of A level students re-sitting examinations in the hope of achieving a higher grade. Students generally welcomed the opportunity for re-sits, with the majority believing re-sits should be allowed in any circumstance. The statistical data indicates, however, that re-sitting may often not result in a better grade outcome.

Stakeholders' perceptions of the impact of changes to GCSEs on teaching and learning

The evaluation has had to consider the four GCSEs with specification changes in 2009 (French, geography, history and Spanish) separately from changes to English and mathematics (changed for first teaching in 2010). In part, this is because of the different timing of their implementation and the degree to which changes to English and mathematics have 'bedded in' compared with the other GCSEs; but it is also because English and mathematics are high-stakes qualifications (as a result of school 'floor-targets') and 'gate-keeper' qualifications for student progression. Assessment of English and mathematics, therefore, is usually the focus of intensive monitoring and a range of strategies for ensuring students reach the required grade. This is not to suggest, however, that a similarly robust approach is never adopted in relation to other subjects.

Student engagement

Although the updating of course content is reported to have increased students' sense of motivation and their engagement with GCSE subjects, this positive effect has been somewhat counteracted by the increased focus in teaching and learning on assessment. With English, there is general concern among teachers that the amount of content is leading to less depth of study and that, in some instances, students are not being required to read a whole text. And with mathematics, there is a range of opinions on the degree of stretch and challenge for students of differing abilities.

Unitised assessment

There were some concerns about unitised assessment for GCSE, but again views were mixed. Some thought that linear assessment promoted more in-depth and longer-lasting learning than unitised assessment, particularly in relation to the development of subject-specific skills; others liked to have the option of unitised assessment for students who learned better by having the opportunity to achieve along the way and to build on previous results.

Some stakeholders considered that unitised assessment, with its opportunities for re-sits and early entry, coupled with pressure on centres and students to get results, was at odds with the value placed on synoptic assessment and on the development of skills rather than a narrow knowledge-based approach to subjects. The overall consensus from centres and awarding organisations was that there is conflict between the need to gain the 'results' required for school performance targets – using re-sits and early entry to maximise pass rates (e.g. at A*–C at GCSE), which is possible with unitised qualifications – and encouraging learning, development, coherence and a greater understanding of the subject.

Controlled assessment

A wide range of approaches to controlled assessment[10] is being implemented within and across centres, often depending on the subject being assessed and/or on the teacher. Awarding organisations have often interpreted Ofqual's regulations differently, so that there is sometimes variation in guidance for the same subject across the different awarding organisations. This has resulted in considerable variation in the amount of support that teachers believe they can give their students in the preparation stage of controlled assessment.

[10] Controlled assessment is a new form of internal assessment of the work of a course, replacing coursework. There is no controlled assessment for mathematics.

Concern was expressed that some teachers are using strategies, even if often unwillingly, to help students to get the best grade, and that these undermine the validity of the controlled assessments in terms of the knowledge and skills that are meant to be assessed. Examples of strategies include running practice assessments that change little in the 'real' assessments and modern foreign language teachers setting more controlled assessments than required and choosing the best for each individual student.

Most centres do not feel that the management of controlled assessment is a problem, although this was not always the case and it continues to be an issue for some centres more than others. About half of the centres to have previously found that the scheduling of controlled assessments involved a significant added management burden now reported that they were coping well or satisfactorily with the issue. The remaining half thought that the challenge of fitting controlled assessments into the school calendar was becoming a more serious issue with the increased number of departments involved. Staff at all levels expressed the need for guidance on best practice in the management of controlled assessment.

It is accepted that the introduction of controlled assessment is one of a number of changes that may increase teachers' workloads initially, but usually only in the short term. Teachers of modern foreign languages, however, reported that they expect their workloads to remain heavier in the longer term because of specific requirements to do with organising and conducting controlled assessment.

Most teachers liked the opportunity that controlled assessment gives to choose topics and contextualise the tasks for their students. However, there was no consensus across or within centres and subject areas about the extent to which controlled assessment was an appropriate form of assessment for particular skills and knowledge. There were concerns that, rather than promoting in-depth independent learning, controlled assessment may only test the ability to learn content and regurgitate it.

Students of history and geography were the most positive about controlled assessment, students of Spanish and French the most negative. History and geography students liked being able to use their research notes during the controlled assessment task. The majority of students also stated that they liked the fact that controlled assessment did not take place in the main school hall, where they sat external examinations, and that it was therefore less stressful.

There is evidence from many of the case-study centres to suggest that the focus on school performance measures is a strong driver for centres to ensure students meet target grades. In many cases this means that teaching and learning is driven by the assessment regime.

There is evidence to suggest that, unless specific content and/or skills are expected to be assessed, some teachers and students will consider them not so much a priority as an indulgence. In some centres, however, the focus is less narrow, with centres embracing the change and challenge presented.

Choice of specification was reportedly determined by several factors, including the appropriateness for the needs and aspirations of students, consistency with the nature of the subject, and familiarity with the awarding organisation.

Grade outcomes at GCSE

As for A level, two important caveats need to be placed on the findings for GCSE:

- summer 2011 data is for entries rather than candidates, which may mean that the effects of re-taking (the entire qualification) cloud information about overall outcomes
- no data is available yet for candidate characteristics, apart from gender.

Nevertheless, on the basis that there is no reason to suspect either major changes in re-sitting behaviour (in GCSE qualifications, not units) since 2010, or significant differences between the 2011 Year 11 cohort and the 2010 and previous cohorts in terms of trends in prior attainment or other important characteristics, it is possible to draw some tentative findings:

- The overall grade distributions obtained in the new specification GCSEs in 2011 are worse than those obtained in 2010, a slight dip following steadily increasing grades since 2004. While the outcomes for females plateaued in 2011 compared with 2010, the outcomes for males fell (by around 0.06 of a grade on average).
- Overall grade distributions have plateaued for all candidates in GCSE French and geography, and decreased a little for GCSE history and Spanish.

The overall grade distributions for male entries have got significantly worse for all four new-specification GCSE subjects under consideration: French, geography, history and Spanish. These changes are not seen for English and mathematics, for which specifications had not changed in summer 2011. In addition, for the four new-specification subjects, the gap between girls' and boys' grades (girls performing consistently better over time in each subject) widened in 2011. This suggests that:

- The new GCSE specifications are proving a little more challenging for boys than girls, even taking into account attainment differences in previous years of GCSE.
- A similar plateauing effect (slightly more pronounced, in fact – a small dip in GCSE grades in 2011) is observed to that seen for A level in 2010, suggesting

that Ofqual's revised guidance to awarding organisations has had the effect of maintaining similar overall grade distributions in a situation where grades might otherwise have been expected to get worse. It should be noted that this last conclusion can be confirmed only by more-detailed analysis of a wider range of new-specification qualifications, coupled with consideration of candidates' prior attainment for the years before and after the specification change.

1 Introduction

This report presents findings from the independent evaluation of the impact of changes to A levels and GCSEs in England. At A level, these changes were introduced to all subjects, except mathematics, in September 2008. The changes included a move from six units to four for the majority of subjects, the introduction of greater stretch and challenge at A2,[11] and the introduction of the A* grade.[12] From September 2009, changes to GCSEs were introduced in all subjects except English, mathematics, ICT (information, communication and technology) and the sciences. Changes to GCSE English, mathematics and ICT were introduced in September 2010. At GCSE, controlled assessment[13] and unitised assessment were widely introduced.

The report follows three rounds of fieldwork with case-study centres and wider stakeholder groups (see Table 1) designed to ascertain their understanding and perceptions of the changes, and their attitudes towards them. It builds on findings from the statistical data reported on in spring 2011,[14] and reports on early indications of the impact of the changes, and of the issues that have arisen.

[11] Awarding organisations were asked to use a variety of stems in questions – for example analyse, evaluate, discuss, compare – to elicit a full range of response types and thereby avoid a formulaic approach; ensure connectivity between sections of questions, thereby avoiding questions that are too atomistic; develop questions that require extended writing in all subjects, except where inappropriate, for example, in mathematics; use a wider range of question type to address different skills, i.e. not just short answer and structured questions, but open-ended questions, case studies, etc; and improve synoptic assessment.

[12] The A* grade was awarded for the first time in summer 2010 to candidates who achieved a grade A on the A level overall (80% on the uniform mark scale), and who also achieved at least 90% or more across the A2 units.

[13] The purpose of controlled assessment is to assess those aspects of a subject that cannot be easily assessed by external examination. These aspects include research, planning, investigation, analysis, collaborative working, and presenting ideas and arguments supported by evidence. Controlled assessment is designed to encourage a more integrated approach to teaching, learning and assessment, and to enable teachers to confirm that students carry out the work involved.

[14] DfE (2011) 'The evaluation of the impact of changes to A levels and GCSEs – second interim report', reference DFE-RR170 www.education.gov.uk/publications/RSG/NewRsgPublications/Page1/DFE-RR170

Table 1: The different phases of this evaluation study

	Phase of implementation	Evaluation report
Phase 1: Spring/summer 2010	(i) End of a full cycle of A2, and two cycles of AS, but before the summer 2010 examinations (ii) GCSE data collected towards the end of the first year of a two-year course – did not include English and mathematics	first interim report
Phase 2: Autumn 2010–summer 2011	(i) First round of awarding for A level (summer 2010) – third AS cohort started course (ii) Start of second year of new-specification GCSE courses (except English and mathematics) (iii) First year of teaching new specification English and mathematics GCSEs	second interim report
Phase 3: Autumn 2011	(i) Second round of awarding for A level (summer 2011) (ii) First round of awarding for new-specification GCSEs – not English and mathematics (summer 2011) (iii) Start of second year of new-specification English and mathematics courses (awarding summer 2012)	Final Report (spring 2012)

The evaluation was originally reported on in the first interim report (July 2010).[15] Data collected in 2010 focused on stakeholders' perceptions of the impact of the changes. These perceptions were based on an emerging picture. The findings offered evidence-based insights into the initial and short-term impact of the changes on centres, students, awarding organisations and wider stakeholder groups. The second round of data collection took place in autumn 2010 and spring/summer 2011, revisiting initial perceptions, and included analysis

[15] The independent evaluation of the impact of the changes to A levels and GCSEs was commissioned by QCDA in March 2010. The contract was transferred to the DfE in January 2011. The executive summary from the first interim report is published as an appendix to the second interim report: DfE (2011) *The evaluation of the impact of changes to A levels and GCSEs – second interim report*, reference DFE-RR17 www.education.gov.uk/publications/RSG/NewRsgPublications/Page1/DFE-RR170

of statistical attainment data for the first A level results for the new-specification assessments.

This final report follows a full cycle of both qualifications (with the exception of GCSE English and mathematics) and considers findings to date, together with statistical attainment data from the summer 2011 examination series and a further round of interviews with case-study centres and stakeholders.

1.1 Background to the evaluation

The initial round of data collection during 2010 concentrated on six GCSE subjects (French, geography, health and social care, history, media studies and Spanish – for which new specifications were introduced in September 2009) and on six A level subjects (English literature, French, geography, media studies, physics and psychology). A change of focus[16] following the Department for Education's (DfE) take-over of the contract for the evaluation has meant that, at GCSE, data was collected for English/English language and mathematics (for which new specifications were introduced from September 2010) instead of for health and social care and media studies. At A level, data was collected for history instead of media studies.

1.2 Aims of the evaluation

The evaluation has addressed the following key themes:

- stakeholders' perceptions of the new A levels and GCSEs
- the impact of the changes on teaching and learning
- the impact of the changes on centre behaviour and management of change.

Questions addressed under each theme can be found in Appendix 1.

This report focuses on building an understanding of the impact of the changes to specifications for A levels and GCSEs across the different phases of implementation. The specifications have been implemented at incremental points within a two-year timeframe and interpretation of the data may therefore depend on where centres are located on the change continuum for each qualification and/or subject.

The report follows a full cycle of both qualifications (with the exception of GCSE English and mathematics) and considers findings to date, together with statistical attainment data from

[16] English and mathematics, the new GCSEs added to the evaluation, are included in the national curriculum for Key Stage 4 (2007). The change of focus at GCSE and A level was made at the request of the DfE, to reflect the coalition government's interest in the 'core' subjects.

the summer 2011 examination series and a further round of interviews with case-study centres and stakeholders.

The report charts the process of change and the impact seen in centres during the evaluation period (March 2010–December 2012).

1.3 Data sources

The following data sources have been used for the evaluation

- Centre research study (CReSt) data (provided by QCDA from the CReSt project) 2009
- Literature review: the Qualifications and Curriculum Development Agency's (QCDA) Research Evidence Management System (REMS) database (undertaken in March 2010)
- Awarding body data archive (ABDA) – awarding organisation data supplied by the Office of Qualifications and Examinations (Ofqual) Regulation relating to A level unit results and re-sitting data, not sample candidate scripts – which are also collected under the ABDA programme. The project considered A level data for English literature, French, geography, media studies, physics and psychology for the summer 2008 and 2010 series (ABDA data is collected on alternate years for GCSE and GCE, starting in 2008). For GCSE, the project considered French, geography, health and social care, history, media studies and Spanish data for the summer 2009 series (the data for summer 2011 does not become available until March 2012 approximately). Note that ABDA data is included only for certain strands/specifications of A–levels, generally the most popular ones in terms of entry volumes.
- Official statistics – statistical first releases (SFR) – from the DfE Research and Statistics Gateway
- Joint Council for Qualifications (JCQ) data
- National Pupil Database (NPD) – including census, pupil level annual school census (PLASC) and examination results
- data from 17 case-study centres – including interviews or focus groups with senior leadership teams, heads of department, subject teachers, examination officers, and students

- interviews with 19 wider stakeholder representative groups[17] and 35 responses to an online survey from 29 English higher education institutions (HEIs)[18]
- awarding organisation and Ofqual document review (updated in 2011 to include new subjects)[19]
- centre online survey data – 44 centres responded in total (6.4% response rate): 19 completed responses to the senior leader questionnaire, 97 individual responses from 41 different centres to the subject-specific questionnaires.

1.4 Scope and limitations of this report

This report includes data on A levels and GCSEs (excluding half-GCSEs and Applied GCSEs). The original round of data collection (commissioned by QCDA) included further education (FE) centres as wider stakeholders. In the second round of data collection (following transfer of the evaluation to the DfE), FE centres were included as providers of full-time A level provision for 16–18 year olds. However, the scope of the evaluation does not include any additional statistical datasets specific to this cohort, as FE attainment data has not been identified as a priority for the resource available.

2 Sources of evidence

2.1 Centre-based case-study data

All except two of the 17 case-study centres originally visited in May/June 2010 were revisited in spring 2011; the other two centres were replaced with an FE college and a secondary school (both of which had sixth forms). Details on how the case-study sample was selected can be found in Appendix 2, together with a breakdown of centre characteristics for the second round of data collection. Overall, there were visits to 17 case-study sites. The visits included interviews with senior leadership teams, heads of department, subject teachers, examination officers and student focus groups.

[17] Employers, subject and professional associations, FE representative organisations, trade unions, workforce development agencies and awarding organisations

[18] An invitation to take part in an online survey was sent to 103 HEIs in England and followed up with reminder emails and telephone calls; Supporting Professionalism in Admissions (SPA) also promoted the survey at one of their events and in a newsletter. Response rate of 27% from the original 103 HEIs invited – one additional HEI also responded.

[19] The review of the awarding organisation documentation follows the logical development process of the revised qualification: the original qualification criteria and the revised subject criteria for each qualification, and the responding awarding organisation specification(s); the sample assessment materials, along with relevant mark schemes, and any other qualification-specific awarding organisation guidance material aimed at teachers and learners [note that the guidance material is not part of the accreditation process].

In autumn 2011 follow-up telephone interviews were undertaken with 10 of the case-study centres. To ensure that there was a greater breadth of data, a further seven centres were contacted and asked to take part in shorter telephone interviews on the same topics as the case-study centres. A total of 33 interviews were conducted with heads of Key Stage 4 (KS4) and KS5, and subject leaders for GCSE English, GCSE mathematics and A level physics. The specific focus on GCSE mathematics and GCSE English/English language and English literature this time was because the new specifications in these subjects were introduced for first teaching in September 2010 only – other subjects were included in more detail in previous rounds of data collection. In addition, because centres had previously reported concern about a lack of stretch and challenge in A level physics, further exploration of the issues was investigated for this report.

Details of the subjects taught by the staff interviewed across the two visits and the telephone interviews, and of their role, are given in tables in Appendix 2.

2.2 Wider stakeholder data

2.2.1 Interviews with stakeholders

The activity reported here straddles the three rounds of data collection. Stakeholder telephone interviews or face-to-face group interviews were held with an agreed sample from the following groups: employer organisations, subject and professional associations, FE representative organisations, trade unions, workforce development agencies and awarding organisations.

The findings were drawn from the analysis of data from interviews with a total of 19 wider stakeholder organisations.

2.2.2 Higher education institution survey

An invitation to take part in an online survey was sent to 103 HEIs in England. The survey ran from 23 March 2011 to 27 May 2011. Overall, 35 responses were received from 29 HEIs (seven from different departments at one university).[20] Of the 29 institutions responding:

- there were 15 pre-1992 universities and 11 post-1992 universities, with two university colleges and one private provider; 52% of the HEIs in the survey are therefore pre-1992 universities, against a national figure of 49%
- 7 of the 29 HEIs were Russell Group Universities, which means that Russell Group Universities are slightly over-represented in the survey (24% of survey

[20] Response rate of 27% from the original 103 HEIs invited – one additional HEI also responded.

HEIs against 15% nationally – there are 16 English members of the Russell Group).

The response rate means that findings should not be generalised to the whole population of HEIs.

2.3 National datasets used

The work undertaken in summer 2010 aimed to establish a baseline dataset for results that pre-dated the first results from the new-specification A levels. The second interim report published winter 2011) looked at the impact of the new specifications on participation, progression and attainment on the basis of the first full suite of results. Those investigations focused on change (expected and unexpected within those three main measures). For the previous cycle of analysis, the investigation was based on the following datasets:

- **ABDA data provided by Ofqual (previously provided by QCDA)** – this includes GCSE and A level exam results from KS4 and KS5. This data, which was anonymised, included unit performance, prior attainment and re-sit information. Starting in 2008, with GCSE and A level, results have been collected in alternate years, for only a sample of subjects, and for only a single specification/specification strand within each awarding organisation's offering for that subject (usually the specification with the highest candidate volume in cases where awarding organisations provide more than one specification for a subject). This subject selection aligns closely, however, with the sample to be investigated in the evaluation. Across the subjects considered in this report, the ABDA data accounts for around 60% of all candidates taking the subjects. This data is identified as 'ABDA' in the report.
- **Official statistics: DfE Research Gateway SFR**[21] – these datasets provide aggregate entry and achievement data for individual subjects (consolidated for all awarding organisations) and for the key school targets (e.g. five A*–C grades including English and mathematics) for candidates completing KS4 and KS5. They are essentially pre-run reports on the NPD dataset but are considered separately here, because the quality assurance of those reports is undertaken by the DfE rather than the evaluation team. This data is identified as 'official statistics' or 'SFR' in the report.

[21] From the DfE Research Gateway SFR data: http://www.education.gov.uk/rsgateway/index.shtml

- **JCQ**[22] – the JCQ provide high-level aggregate summaries of the candidature (including by gender) and profile of grades awarded, which is useful because it is the first dataset published (in August) after the release of new results.
- **DfE NPD** – the DfE provided the evaluation team with a subset of these large datasets, including Key Stage 3 (KS3), KS4 and KS5 results and a range of demographic indicators about students and the schools they attend. The datasets provide identifiable candidate data, thereby allowing matching across the datasets for progression analysis. Data from the DfE national pupil database is referred to as 'NPD' data in the report.

For the final stage of statistical analysis data (summer 2011 examination results) further data is available from provisional SFR data and JCQ data only.

2.3.1 AS level participation data

The data presented here excludes AS level information for attainment and participation presented in the SFR dataset because the AS level candidature reported in SFR is known to under-report the actual number of candidates taking AS level qualifications – some candidates record no attainment in Year 12 because they do not cash in their AS level results until Year 13, at the time they submit for the A level qualification. Official statistics and NPD data record only the 'highest' qualification in a subject for a candidate in any given year, so, when a candidate cashes in their AS results at the same time as their A level qualification (rather than having cashed it in during the previous year), the AS award is not recorded at all in official statistics, either in Year 11 or 12 results. This situation has changed, starting with the 2011 summer series, where candidates in all maintained centres in England are required to cash in their AS results in Year 12.

This finding was reported in outline in the second interim report[23] and has been investigated in a little more detail for this report, with the result that more-accurate rates of progression from AS level to A level can be reported for the six subjects under consideration. This work has been undertaken on the basis of NPD data and is presented in Section 3.5.

2.3.2 Calculation of mean grade score for A level

The mean grade scores presented here combine the proportion of candidates achieving each grade into a single grade score by assigning a weight to each grade and multiplying that weight by the proportion of students achieving the grade. For the calculations made, an

[22] http://www.jcq.org.uk/national_results/index.cfm

[23] DfE (2011) The *evaluation of the impact of changes to A levels and GCSEs – second interim report*, reference DFE-RR170, November 2011 (for example page 30) www.education.gov.uk/publications/RSG/NewRsgPublications/Page1/DFE-RR170

A* or A grade is weighted as 5, B as 4, C as 3, D as 2 and E as 1. All other grades (i.e. U and X) are given zero weighting. Since the A* grade was introduced in results only from 2010 onwards, it is given equal weighting to an A grade so that fair comparisons may be made (on the basis that candidates receiving an A* grade would have been previously awarded an A). A high mean grade score would suggest the pupils did well overall, while a low mean grade score would suggest the candidates did poorly overall.

Of particular interest in this report is the performance in the second year of the new-specification A and AS levels compared with performance in previous years on the old specifications. In 2010, a slight plateauing in the grade profiles was seen in most subjects when compared with the rising trend of previous years, and it will be of interest to see whether this continues in future years or returns to the previous trend. Comparing mean grade scores calculated from the time series data available enables such changes in trend to be identified, if there are any.

2.3.3 Calculation of GCSE mean grade score

The mean grade scores presented here combine the proportion of candidates achieving each grade into a single grade score, calculated by assigning a weight to each grade and multiplying that weight by the proportion of students achieving the grade. For the calculations made, an A* grade is weighted as 8, A as 7, B as 6, C as 5, D as 4, E as 3, F as 2 and G as 1. All other grades (i.e. U and X) are given zero weighting. As a result, a high mean grade score applies to candidates who did well overall, while a low mean grade score applies to candidates who did poorly overall. It should be noted that mean grade score does not provide information about the profile of individual grades, and so, although it is used widely in this report, where the proportions achieving particular grades are concerned (e.g. A* grades at A level), the proportions obtaining that specific grade are shown.[24]

Of particular interest in this report is candidates' performance in the first year of the new-specification GCSEs (the majority of GCSEs completed in summer 2011 – only English, mathematics, ICT and science remained on legacy specifications) compared with performance in previous years on the legacy specifications. Comparing mean grade scores calculated from the time series data available enables any changes in the trends of results to be identified.

[24] For example one set of grades might have 50% of candidates with a C grade and 50% with a D; another set might have 12% with each grade A*–G. The mean grade score would be the same for each, but the actual profile of grades would be very different. This is an impractical example – but the general point applies. If a mean grade score goes up, it could mean that there were proportionately more A* candidates and fewer G candidates (with all other grades staying the same); or it could mean proportionately more of all A*–C candidates and fewer of all D–G.

Although there are non-linear approaches to calculating average grade scores (for example, in points-based scoring for progression such as UCAS application), only a simple linear calculation is used here; assigning non-linear weightings to grades would be inappropriate for this work.

2.3.4 Note on Ofqual guidance to awarding organisations

In the second interim report,[25] the impact of new guidance from Ofqual to awarding organisations about grading processes for the new-specification A level awards in summer 2010 was considered. In summary, Ofqual made explicit the requirement that candidates completing new specifications should not be advantaged or, more likely, disadvantaged as a consequence of being the first group to do so, noting that in previous instances of change candidates had fared less well than predecessors with comparable prior attainment.

The new Ofqual guidance for judgemental grade boundary-setting procedures included more emphasis on consideration of prior attainment data at GCSE, with the requirement that, if the 2010 A level cohort's Year 11 GCSE grades (from 2008) were comparable with those achieved by A level candidates in the previous year, then the grades awarded to the 2010 cohort should be comparable with the previous year, too – and that, if awarding organisations chose not to follow this guidance, they were to explain to Ofqual why they had not.

These new guidelines about setting the overall standard on the basis of prior attainment data were probably responsible, at least to some extent, for the plateauing of grades seen in 2010 (an increase in mean grade score of 0.17% from 2009 to 2010 compared with an average of 1.61% year on year from 2003 to 2009). Indeed, it appears that the pattern of A level grades awarded in 2010 might well have been less good than that in previous years, had it not been for the application of this specific Ofqual guidance.

The arrangements made by Ofqual for A levels in 2010[26] were continued for 2011 with the additional concern to avoid the likely increase in outcomes as centres become more familiar

[25] DfE (2011) *The evaluation of the impact of changes to A levels and GCSEs – second interim report*, reference DFE-RR170, pp November 2011, section 3.1.1 pp18–20 www.education.gov.uk/publications/RSG/NewRsgPublications/Page1/DFE-RR170.

[26] The prime objectives of maintaining grade standards over time and across different specifications within a qualification type (Ofqual, April 2010, 'GCSE, GCE, Principal Learning and Project Code of Practice', para 6.22) necessarily become more problematic, and engender more concerns among stakeholders, at times of curricular change. Noting this, Ofqual issued new guidance to awarding organisations for the summer 2010 A level series (the first year of awards at A2 for the new specifications). The guidance itself is unpublished, but information was provided to headteachers and MPs: http://www.ofqual.gov.uk/help-and-support/94-articles/341-changes-to-A-levels-in-summer-2010, May 2010, etc, retrieved 19 January 2012.

with the new specifications.[27] Ofqual has also published a statement confirming that similar arrangements were introduced for new-specification GCSE results in 2011, based on prior attainment from Key Stage 2 (KS2) assessments.

2.3.5 Grade outcomes

For each subject of focus at GCSE level, a breakdown of mean grade score by gender and year is provided, as well as a graph indicating the difference in mean grade score between males and females each year.

Previous reports have focused on SFR data,[28] because it relates specifically to KS4 students and can be more easily tied to the NPD datasets. However, a main area of interest in this year's results is the gender difference, for which JCQ data[29] provides the most detailed information (and the data reported in SFR for 2011 is not directly comparable with that of previous years).[30] The JCQ data also ensures that achievements relate to the new-specification qualifications rather than to the legacy qualifications that may have been taken in earlier years (in the case of SFR). Work in previous reports has shown that the majority of GCSE candidates are of school age, so it is reasonable to infer that gender and grade-awarding information from JCQ data can be interpreted as relating to school-age candidates.

It should be noted that the data is based on entries rather than candidates. It is possible, therefore, that changes in grade profile may relate to candidates' taking more attempts at a particular subject with a consequently higher proportion of poor grades, rather than reveal any trend in overall outcomes at the end of the learning programme. Although this can be resolved for certain only when NPD data is available for analysis, qualitative evidence from centre interviews suggests no major change in patterns of qualification re-sitting.

The A level attainment and entries by subject for candidate entries in England are also reported here, using data from SFRs, as published on the DfE Research and Statistics Gateway, since it is students of school age (16–18) who are of most interest in this report.

---

[27] In 2011, the second year of new specification A2 awards, the arrangements introduced for 2010 were continued with the additional purpose stated. The arrangements were also applied to GCSEs for the summer 2011 series, the first year of completions of the new specifications. As before, the guidance is unpublished but information was provided to headteachers: http://www.ofqual.gov.uk/files/29-03-11-openletter-schools-colleges-gcsesummerawards.pdf, 14 March 2011, retrieved 19 January 2012. Further helpful information is included in Item 24/11 of Ofqual's Board, dated 8 June 2011, here, retrieved 19 January 2012.

[28] DfE SFR, provisional, 20 October 2011, http://www.education.gov.uk/rsgateway/DB/SFR/s001034/index.shtml, retrieved 27-11-11, and revised releases for previous years, e.g. http://www.education.gov.uk/rsgateway/DB/SFR/s000985/index.shtml.

[29] JCQ, summer 2011, etc http://www.jcq.org.uk/national_results/gcses/, retrieved 27-11-11.

[30] In the 2011 provisional SFR, Table 11 reports the best grade achieved by each candidate during their time to the end of KS4 rather than recording all entries and grades achieved, as was the case in 2010 and before.

This provides A and AS level attainment and entries for all students aged 16–18 at the beginning of the academic year.

As for the GCSE data, the data shown is for candidate entries in each year for candidates aged 16–18, not for individual candidates. Re-sitting of entire qualifications at A and AS level is relatively uncommon, however, so the results may be treated as a reasonable proxy for candidate outcomes, taking note of the caveats raised in the sections on GCSE above.

Because the SFR dataset includes proportions of candidates achieving each of the available grades, by gender, this data has been used in favour of JCQ data (in contrast to the GCSE analysis, where JCQ data is used, as it includes such information).

It should be noted that, in contrast to previous reports, no information is available in this report about, among other things, candidates' prior attainment, profiles of re-sits, range of qualifications taken, personal characteristics and demographics. This information is not released in the NPD datasets until March of the year following the August results, and is available in the ABDA data collated by Ofqual for research purposes only around the same time.

Together, these factors place substantial limitations on the analysis. Although it is possible to identify trends in attainment, it is not possible to determine whether these relate to changes in the students taking the examinations, to the examinations themselves or to other factors. As a result, the quantitative analysis in this report has a number of caveats associated with it.

2.4 Centre online survey

The centre survey was delivered as nine separate online questionnaires. The survey was available from 10 May 2011 to 10 June 2011. The nine questionnaires were:

- four combined A level and GCSE questionnaires – English, French, geography, history
- two A level-only questionnaires – psychology, physics
- two GCSE-only questionnaires – mathematics, Spanish
- one generic senior leaders' questionnaire.

Most of the questions posed in the eight subject-specific questionnaires were essentially the same for each subject/level but, where appropriate, the questions were specifically worded for the subject. The questionnaires for A level only and for GCSE only were essentially subsets of the combined A level and GCSE questionnaires. The questions generated quantitative (nominal, ordinal and ratio) and qualitative data.

A total of 691 centres were invited to take part in the survey, of which 44 responded to one questionnaire or more.[31] Centres that did not respond cited the timing (pre-summer examinations) and existing heavy workloads as reasons for not doing so. An overview of centre characteristics for participating centres is included in Appendix 2.2. Although the low level of response (6.4%) to the online survey means that any findings should be treated with caution, and no general conclusions may be drawn from them, they have nevertheless been included, where appropriate, to validate or challenge findings from the case-study data.

3 Research findings

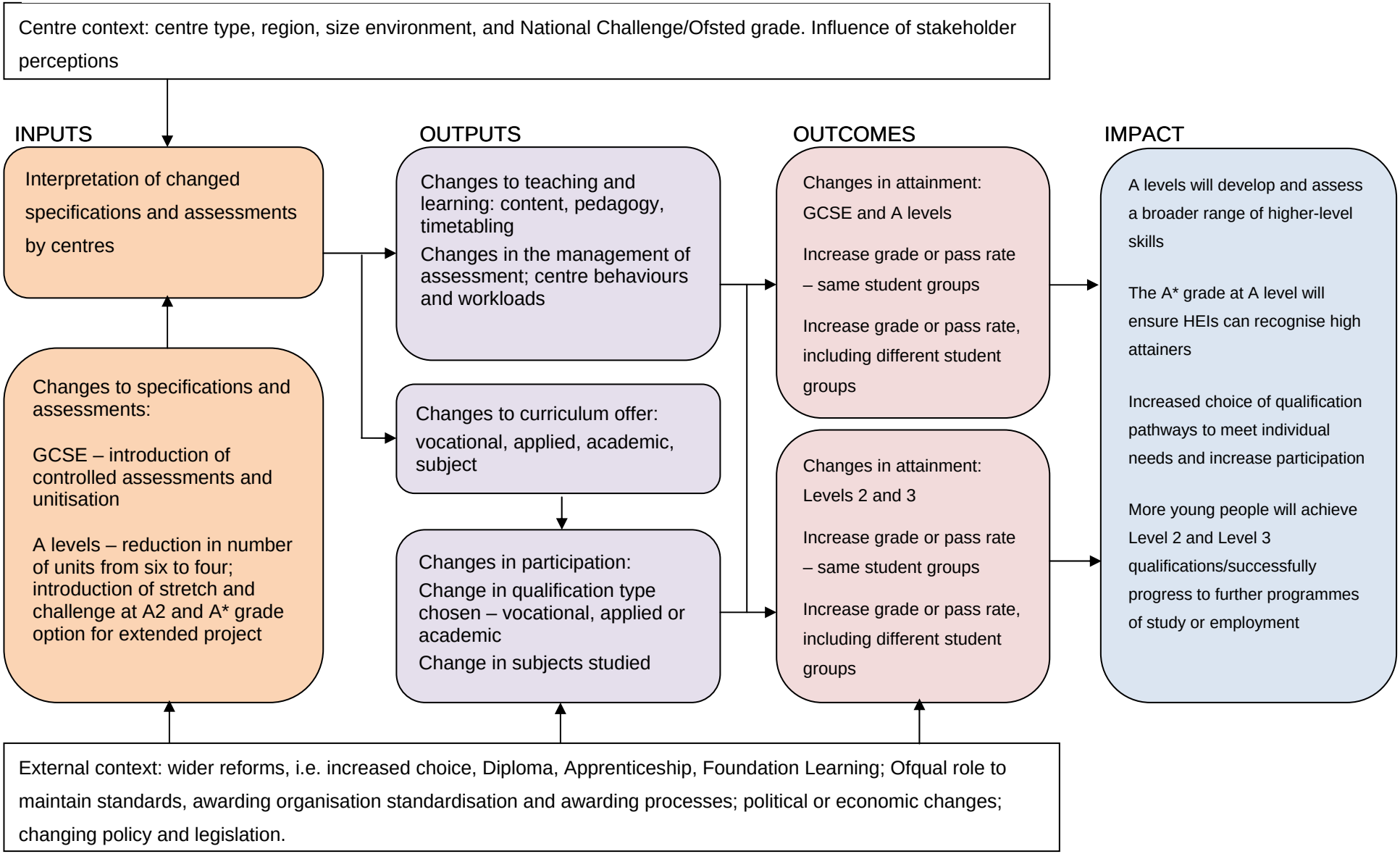
3.1 Wider context

3.1.1 Overview of literature reviewed to date

The original 'theory-of-change' model developed prior to the first round of data collection in spring 2010 outlined the drivers and influences on the potential impact of the changes to A level and GCSE specifications (Figure 1). The theory of change was developed on the basis of the initial literature and document review (March 2010).

[31] An initial stratified sample was identified from Edubase (whole population listed as 4,007 centres, although some of these were later removed after further research had identified them as inappropriate – because they were consortia, for example, or adult education centres, or because the centre had closed). Further centres were identified using a purposive strategy if there was a gap in the data available via Edubase (as was the case with FE colleges). Strata used were transition (to age 16, through 16, 16+), prior attainment (selective, non-selective), level of deprivation (calculated on the basis of the percentage of students entitled to free school meals), teaching and learning (in special measures, not in special measures).

Figure 1: Theory of change v1: intended impact of changes to GCSEs and A levels based on the literature review



The changes to A levels and GCSEs were part of wider curriculum reforms in response to the 14–19 Education and Skills White Paper (2005).

The aims of the 14–19 Education and Skills White Paper relating to GCSEs were to:

- review coursework with the aim of reducing the assessment burden
- review mathematics GCSE qualifications with the aim of improving motivation and progression to A level in line with recommendations by Professor Adrian Smith (2004)[32] – this was expected to include the linked pair of GCSEs in mathematics
- offer new science GCSEs which would support the aim that young people should do two science GCSEs.

The aims for A levels were to:

- introduce stretch and challenge within A levels through the introduction of Advanced Extension Award-style questions and the 'extended project', with the aim of stretching young people and assessing a wider range of higher-level skills
- reduce the assessment burden by reducing the number of units from six to four but without any change in the overall content of A levels
- ensure there are natural progression routes both through the levels of the Diploma and between GCSEs and A levels and the different levels of the Diploma to offer routes that avoid early narrowing down, and instead give a choice of what to learn and in what setting
- ensure universities have more information on which to make judgements about candidates, by ensuring that they have access to the grades achieved by young people in individual modules.

These aims were translated as part of the wider reforms within the context of the 'knowledge' economy and the need for academic and technical skills, complemented by a wider range of workplace competencies and softer skills. Reform of 14–19 education and training was seen as part of this bigger picture for education.

The literature reviewed in March 2010 indicated a range of other influences on student participation, attainment and progression, relevant at KS4 and KS5, including:

[32] Smith, A. (2004) *Making Mathematics Count: the report of Professor Adrian Smith's inquiry into post-14 mathematics education.* London: The Stationery Office.

- the level of teacher interest and enthusiasm, the ability to make teaching and learning engaging and the level of individual help students received[33]
- factors relating to outstanding teaching and learning at Level 3 based on teachers' good knowledge of students' prior attainment, regular assessment of performance and an understanding of personal circumstances that might affect students' learning[34]
- the impact of socio-economic factors on student attainment and aspirations[35]
- the extent to which the education and training system creates and sustains, for each student, positive attitudes towards the act of learning itself and towards progression, the incentives created in wider society and within the labour market, and the rewards they give rise to that are external to the learning process itself.[36]

The impact of these factors, together with policy decisions made since the change of government in May 2010, needs to be considered in evaluating both the extent to which the original aims for A levels and GCSEs have been realised and also what can be learned to influence curriculum and policy development in a changed policy context. The expected outputs and outcomes outlined in Figure 1 (the increased focus on 'core' subjects, the introduction of the English Baccalaureate as a qualification and school performance measure, an increase to 50% in the floor-target for schools for student attainment of 5 A*–C GCSEs, changes to the vocational curriculum – including the cancellation of Phase 4 of the development of the Diploma, the rise in the ceiling placed on future university fees, the further review of the national curriculum, and a move to linear-only assessment) may be seen to be of greater influence on centre and student behaviour than specification changes and need to be considered in any explanation of the findings from the data.

In the report below, the tension between maintaining qualification standards over time (so that students, although potentially experiencing a different challenge in an examination with a new specification, should not be rewarded differently)[37] and the influence of assessment on teaching and learning – such as Popham's (1987)[38] 'measurement-driven instruction' or

[33] Gorard *et al* (2009) *14–19 Reforms: QCA Centre Research Study, commentary on the baseline of evidence 2007–2008* and Ofsted (2008) *A comparison of the effectiveness of level 3 provision in 25 post-16 providers: how well do students achieve on level 3 courses in different post-16 providers and what factors contribute to their achievement?*'(070167).

[34] Ofsted (2008) *A comparison of the effectiveness of level 3 provision in 25 post-16 providers: how well do students achieve on level 3 courses in different post-16 providers and what factors contribute to their achievement?* (070167).

[35] Gorard *et al* (2009) *14–19 Reforms: QCA Centre Research Study, commentary on the baseline of evidence 2007–2008.*

36 Keep, E. (2009 *Internal and External Incentives to Engage in Education and Training – a Framework for Analysing the Forces Acting on Individuals*, SKOPE Monograph No 12, June, Oxford: SKOPE.

[37] See DfES (2002) *Inquiry into A level Standards: Final Report* – the Tomlinson Report, para 22.

[38] Popham, W.J. (1987) The merits of measurement-driven instruction. Phi Delta Kappan 68: 679–82.

Cheng and Curtis's (2004)[39] 'backwash'[40] factor – add to the complex task of identifying the impact of specification change within a changing context.

3.2 A level findings

3.2.1 Attainment

Attainment has been considered here in its widest form: grade outcomes and student development.

3.2.1.1 Overview of statistical data analysis

Figures 2–4 show consolidated grades for all subjects excluding mathematics and further mathematics, the specifications for which did not change in 2010. Following the plateau observed in 2010, grades have resumed an upward trend. Notably, the proportion of candidates obtaining the highest grades (A* and A) has remained the same (24.6%) since 2009, the last year of awards of the old specification. The increase seen in 2011 (compared with 2010) results from increases in the proportion of candidates obtaining grades B and C.[41] Females continue to perform better than males, although the gap, closing slowly from 0.25 of a grade in 2003, has closed to 0.18 in 2011 – and the introduction of the new specification A levels appears to have had no effect on this. There are also anomalies in some subjects, as identified below.

[39] Cheng, L., and A. Curtis (2004) 'Washback or backwash: a review of the impact of testing on teaching and learning' in L. Cheng, L., Watanabe, Y. and Curtis, A. *Washback in Language Testing: Research Contexts and Methods*, ed., 3–17. Mahwah, NJ: Lawrence Erlbaum.

[40] The effects of tests on teaching and learning.

[41] When grade boundaries are set at A level, human judgement is involved only for setting grades at A and E, and the judgement appears to have been to retain the same proportion of candidates achieving grade A (and the A* grade in the case of the 2010 and 2011 series) at the same level (the proportion of candidates awarded the E grade has continued to decline). This suggests that the profile of candidate performance at the intermediate grades (which are calculated arithmetically once the judgemental grades of A and E have been decided) may have changed a little with the introduction of the new specifications, perhaps as a result of the effect of stretch and challenge on scores, for example. It is not possible at this stage to investigate further.

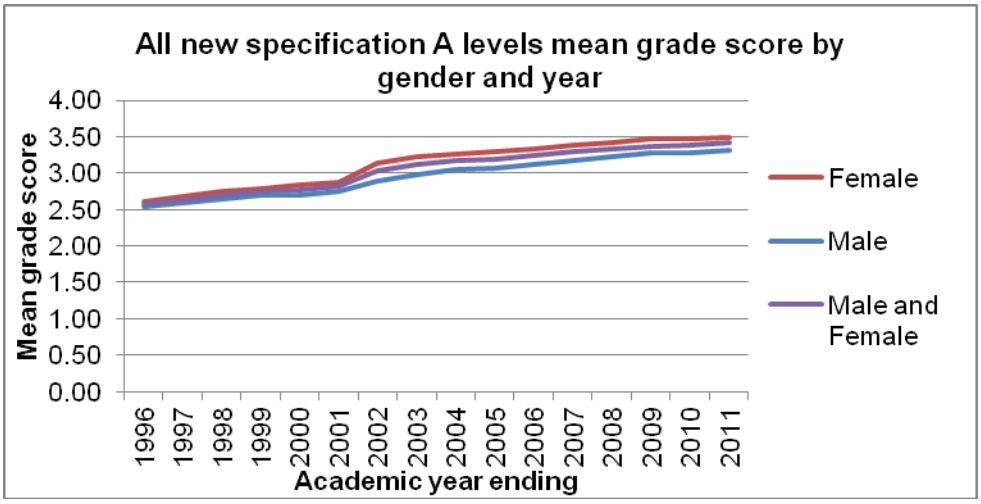


Figure 2: All new-specification A level subjects mean grade score, 1996–2011 (SFR)[42]

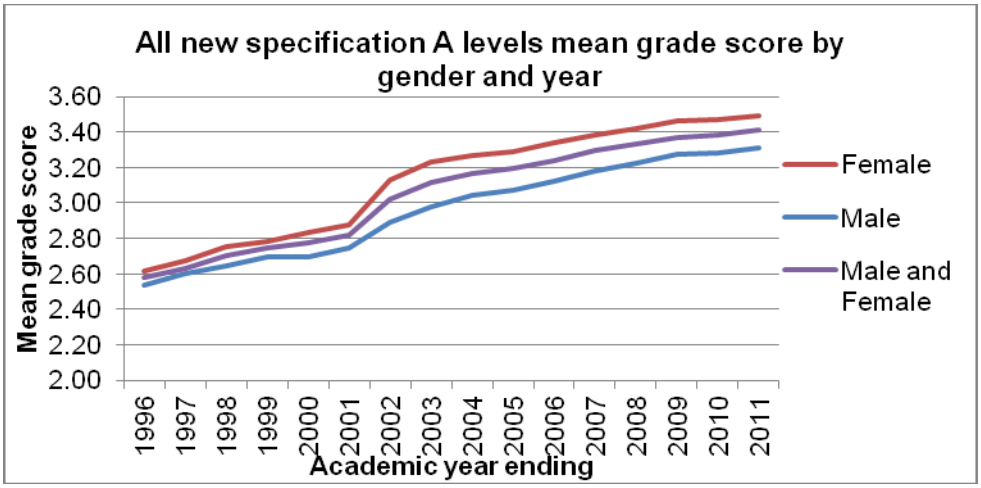


Figure 3: All new specification A level subjects mean grade score, 1996–2011 (SFR) (zoomed y axis – grade score – to show change more clearly)

[42]Figure 2 shows the mean grade score by gender year on year for the new specification A levels. Figure 3 shows only a proportion of the y-axis.

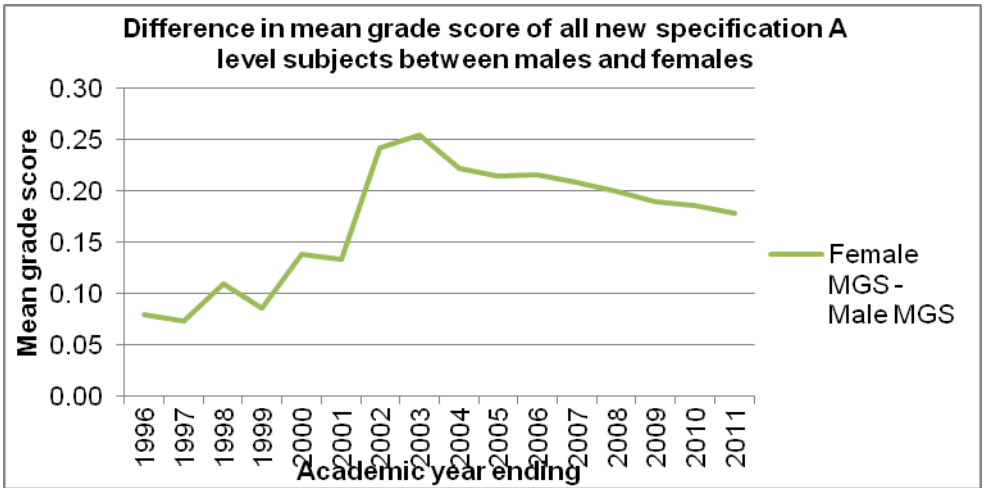


Figure 4: Difference in mean grade score of all new specification A level subjects, by gender, 1996–2011 (SFR)

For each subject of focus at A level, a breakdown of mean grade score by gender and year is provided, based on the SFR data of entries for 16–18 year olds at the start of the academic year (i.e. those in KS5). Also provided for each subject is a graph indicating the difference in mean grade score between males and females each year, also based on the SFR data (Appendix 3.1). Table 2 gives an overview of the change in mean grade score since the change in specification for the six subjects evaluated. It illustrates the plateauing effect seen in 2010, and the increase in grades seen in 2011, as well as showing that, although this is the trend for new specifications as a whole, there are variations from subject to subject.

Table 2: Mean grade score results compared for 2009/10 and 2010/11 (SFR)

Subject at A level	Mean grade score result for 2011 candidatures as a whole (SFR)	Mean grade score result for 2010 candidatures as a whole (SFR)
English literature	No increase of grade points compared with 2010; however, SFR data does not disaggregate for English literature, so this should be used with caution	Decrease of 0.01[43] grade points compared with 2009, following an average increase of 0.035 each year from 2003 to 2009; however, SFR data does not disaggregate for English literature, so this should be used with caution
French	Increase of 0.06 grade points compared with 2010	Increase of 0.05 grade points compared with 2009, following an average increase of 0.05 each year from 2003 to 2009
Geography	Increase of 0.01 grade points compared with 2010	Increase of 0.01 grade points compared with 2009, following an average increase of 0.05 each year from 2003 to 2009
History	Increase of 0.03 grade points compared with 2010	Increase of 0.03 grade points compared with 2009, following an average increase of 0.035 each year from 2003 to 2009
Physics	Increase of 0.03 grade points compared with 2010	Increase of 0.05 grade points compared with 2009, following an average increase of 0.04 each year from 2003 to 2009
Psychology	Decrease of 0.01 grade points compared with 2010	Increase of 0.01 grade points compared with 2009, following an average increase of 0.03 each year from 2003 to 2009
All new-specification subjects	Increase of 0.03 grade points compared with 2010	Increase of 0.01 grade points (the plateauing described earlier) compared with 2009, following an average increase of

[43] A grade point corresponds to a grade, so an increase of one point would mean candidates on average scored one grade higher than the previous year. A decrease of 0.01 is thus very small, but still statistically significant, given the number of candidates completing these qualifications.

Subject at A level	Mean grade score result for 2011 candidatures as a whole (SFR)	Mean grade score result for 2010 candidatures as a whole (SFR)
		0.04 each year from 2003 to 2009

Two important caveats are to be placed on these findings:

- summer 2011 data is for entries rather than candidates, which means that the effects of re-sitting may cloud information about overall outcomes
- no data is available yet for candidate characteristics, so any changes in attainment cannot be attributed either to changes in assessment or to the candidature.

Assuming no reason exists to suggest that the 2011 A level candidate cohort differs significantly from its equivalent in 2010 in terms of prior attainments and other important demographics, or that major changes have occurred in re-sitting behaviour since 2010, the following tentative conclusions may be drawn:

- For new-specification A levels as a whole, the plateau effect seen in the 2010 results appears not to have been repeated in 2011 and the upward trend seen up to 2009 looks to have been largely restored.
- For subjects considered individually, attainment in geography and psychology has continued to plateau, while attainment in French, history and physics has continued to follow an upward trend (even in 2010, little or no plateauing was observed in these three subjects).
- Throughout the two years of the new-specification A levels, the difference in levels of performance between females and males has remained largely unchanged – females consistently achieve higher grades.

3.2.2 Centre perceptions of attainment

Before the publication of examination results in summer 2010, stakeholders had expressed concern that there would be a drop in the overall profile of grades achieved, which suggested that centres did not fully understand or realise that the awarding organisations would maintain grading standards in the face of new challenges presented by revised examination specifications. In the event, the plateau in grades at A level seen in the

statistical data for 2010[44] did not appear to fully reflect the increased level of challenge perceived by students and teachers before the results were known, especially at A2. The most likely explanation for the observed plateau in grades is that it is a direct result of the measures introduced by Ofqual to ensure, first, that the first candidates to complete the new specification A levels were not disadvantaged as a result and, second, (from the 2011 series) to limit any increases in subsequent years as teachers became more familiar with the new specifications.

As recorded in the previous report, analysis of demographics, prior attainment and the impact of re-sitting in the statistical data did not offer an explanation for the overall plateau in A level grades for the new specifications that occurred in summer 2010. Following the examination results for the summer 2010 examinations, case-study centres reported that overall grades at A2 for English literature and geography had been as expected – in other words, had stayed the same or had gone up. For French, history, physics and psychology the centres reported grades had been as expected, stayed the same or gone down. Centres responding to the online survey had also felt that it was more difficult for students to achieve their expected grade: 23 centres said it was harder, 16 that there was no change, and only 3 that it was easier.

Follow-up interviews with centres in autumn 2011 suggest that patterns of student grades overall were either as expected, or better than the previous year. One of the centres reported that results, although not as good as last year, were as expected, given the nature of the cohort. One centre, reporting that summer 2011 results were worse than expected, had concerns about the consistency of marking. The centre had had more than 350 papers re-marked, with every one being awarded a higher grade as a result. Another centre with lower than expected results had also requested that about 20 papers should be remarked, all of which came back with increased marks.

[44] From 1996 to 2009 (the last year in which the previous GCE specifications were awarded), A level grade profiles continued broadly on the same improving trend in most subjects – a small steady increase of about 0.06 of a grade per year in grades attained, punctuated by short-term dips downwards from the trend at times of specification change. Also, since around 2003, participation has been increasing in most subjects.
For results in 2010, a year in which another specification change has occurred (the specification change for teaching from 2008 working through to results in 2010), there was a plateauing of grades following a long period of increasing grades – i.e. another shift downward from the trend, reminiscent of results in previous change periods in curricula (although much smaller in size than the dip in 2000/01). Looking at individual grades attained (across all new-specification subjects combined), the same proportion of students obtained the A grade (including those who received the newly introduced A* grade) in 2010 as received an A grade in 2009 (24.6%), whereas, for example, from 2003 to 2009, on average year on year 0.7% more students received an A grade.

3.3 Depth and breadth of teaching and learning

3.3.1 Stretch and challenge

3.3.1.1 Interpretations of stretch and challenge

Stretch and challenge is being interpreted in a variety of ways. Some centres are uncertain about what is involved, and this was reflected in their responses. These uncertainties have arisen largely as a result of:

- ambiguity about whether the stretch and challenge requirement is intended to be directed at all students, or only the highest attaining
- variations in the ways in which awarding organisations are extending stretch and challenge and applying it in different subjects – greater stretch and challenge may, for example, be introduced:
    - into the qualification syllabus (e.g. increased content, higher-order skills/concepts)
    - by changes to the assessment objectives and/or relative weightings
    - in the paper setting, via amendments to the nature of the questions/tasks
    - in the marking criteria
    - at the grade-awarding process.

Some of these approaches (e.g. amendments to the syllabus content or question paper) are more obvious than others, and centres may then re-focus their teaching and learning appropriately. Tensions arise, however, when centres consider that stretch and challenge in the syllabus (and hence in their approaches to teaching and learning) is not matched, or is not perceived to be matched, by the styles of questioning and marking strategies employed in the assessment. An instance of this occurred when the focus on independent learning was not reflected in examination questions or schemes which, instead, neither asked for, nor credited, independent thought and investigation.

From the online centre survey of individual subjects, of the 59 respondents who expressed a view on the extent of stretch and challenge at AS level:

- 29 thought that students were equally as stretched and challenged as previously
- 24 thought that the changes to the A level specifications had stretched and challenged their students more
- 6 thought that there was less stretch and challenge.

The subject most often perceived as exhibiting greater stretch and challenge than previously was geography, with physics the least changed.

The online survey also showed modest support for the idea that new A level specifications created greater stretch and challenge at A2. Of the 43 that expressed a view, 28 thought that the assessment created greater stretch and challenge, and 25 thought that the content did. Again, the subject for which this view was expressed the least was physics.

Most commonly, centres identified stretch and challenge in terms of the development of independent learning and research skills, the synthesis and conceptualisation of ideas across a subject, depth and breadth of learning, and the demands of the A* grade. Of the eight heads of KS5 interviewed in autumn 2011, over half felt that the new specification A levels stretched and challenged students across all levels of ability. The reasons they gave were broadly based on how they perceived the nature of the examination syllabus – certain subjects, for example mathematics and chemistry, were seen as intrinsically difficult subjects and hence having stretch and challenge necessarily built into them. Alternatively, those subjects which had retained coursework were thought to provide stretch and challenge through the opportunities they offered for students to undertake research and explore the subject as a way of deciding what they wanted to focus on.

Those centres that did not see stretch and challenge as engaging all students were concerned that, where it was built into the examination paper in the form of 'harder' questions, such questions would be fully accessible only to higher-attaining students. Middle-attaining students were liable to 'slip' down a grade, usually because of a weakness in their examination technique or a failure to spot a question's potential for expanding their response. With one centre, however, the extent to which students were being stretched and challenged was considered to rest on the teaching approach adopted by the individual member of staff rather than the syllabus or question paper.

3.3.1.2 Stretch and challenge: the case of physics

In the initial surveys, teachers of physics had appeared unconvinced that the new specifications had achieved the aim of stretching and challenging students. Indeed, at both AS and A2 level, physics was the subject reported as being the least affected by the drive to become more challenging. The further interviews in autumn 2011 sought to understand the reasons for this.

What emerged was a somewhat confusing picture. Only two centres commented that they felt students of all abilities are stretched and challenged by A level physics – and one of these felt that those who are particularly good at mathematics can 'slip through' and not be challenged as much as they might have been, as they can use their mathematical skills to avoid having to conceptualise as much. Broadly, therefore, the consensus among centres was that although middle- and lower-attaining students found the new examination

syllabuses and question papers challenging – as had often previously been the case – they were insufficiently demanding for the more advanced students. This was so much the case that one centre argued that the most able needed to be led beyond the specifications to take on 'significant' extra work.

There was less agreement, however, on the question of why this was the case and how it might be rectified. Among the reasons given were the absence of a synoptic paper, the removal of mathematical-style questions that expect students to use two or three equations, the loss of questions requiring students to describe how they would carry out a particular investigation, and the lack of any demand to memorise formulae. In this latter case, however, the centre went on to acknowledge that in the 'real world' individuals would look these formulae up as the need arose and that the real challenge comes when students have to demonstrate their ability to rearrange and differentiate equations.

Among the amendments that were seen as potentially more stretching, centres argued for a greater emphasis on synoptic learning, practical examinations that really did test thinking skills, and coursework that offered a genuine opportunity for students to take a self-directed learning approach. It was also suggested that greater stretch and challenge would result from more investigative skills assignments (ISA) spread throughout the academic year, giving students the opportunity to perform more practical tasks. This was seen to be more meaningful than the current situation in which students work towards one practical examination.

The view that the specifications were insufficiently challenging for students at the upper end of the ability range carried with it the implication that the reason for this was that the changes had, if anything, helped those lower down the range. But there was, also, reflected in some of the responses a view that physics was a difficult subject, that weaker students struggled – especially with abstract concepts – and that they often found tasks that required application of knowledge to be beyond them. Centres have recognised this, which is why course entry requirements have in many cases become more rigorous. Introducing greater stretch and challenge in ways that would not simply deny many students access to large parts of the curriculum is clearly highly problematic and warrants further exploration.

3.3.2 Independent learning

In the initial case-study visits, the greater emphasis in the specifications on independent learning was generally seen as positive, to be welcomed as furthering both student and staff development. Independent learning stretched and challenged students by encouraging them to work in more exploratory and autonomous ways – particularly where the course content was seen to be no longer content driven, but skills driven. In such instances, teachers felt

that a new emphasis on activities such as decoding research, looking at more-abstract problems, dealing with different conceptual issues and viewpoints – as opposed to learning facts and regurgitating them – had all been positive, albeit challenging, changes to the specifications.

The extended project qualification was felt to be one of the ways in which independent learning, and hence stretch and challenge, was encouraged. A minority of centres, however, expressed concerns about both the extended project in particular and independent learning more generally. On a practical note, they were finding that to implement the extended project properly required intensive staff resourcing, with teachers working individually with each student to get them started:

> *Independence is not 'Here you are, here's something, go away and do it'. Independence is only going to be effective if it's structured, and the extended project, which is I think a fantastic model, actually in reality will take so many hours of staff time that it's likely to bring us to our knees. Well, we're already on our knees. It's likely to see us off. So we're in that awful bind where the best ways of stretch and challenge are just not manageable because there's no way a 16 or 17 year old can do something really meaningful on their own without scaffolding, without staff support.*

Further concerns were that a transition period was needed while students got accustomed to working in a more independent way, and that the overloaded course content at times meant that independent study was very outcome-focused rather than truly independent.

3.3.3 The development of subject-specific skills and knowledge

Data from the visits to the case-study centres in spring/summer 2011 reported the commonly held view across the six A level subjects that the new specifications had brought about positive changes to teaching and learning. The interviews undertaken were with subject heads of department and teachers. In English literature, French and geography, the reduction from six units to four was seen as having introduced greater challenge and depth into teaching and learning, encouraging the development in depth of subject-specific skills and of greater breadth of subject-specific knowledge.

In summer 2010, case-study centre teachers of different subject areas, and those working with different awarding organisations, had appeared to disagree about the impact of the reduction in the number of units. There had been concerns about the amount of content overloading teaching and learning, so that the focus was on input rather than developing higher-order subject-specific and generic skills. In a few case-study centres, teachers reported that they continued to use didactic, essentially transmission-based (rather than

developmentally based) approaches to teaching, because they felt there was so much subject content for students to be given.

In 2010, many teachers had felt that the fact that there was insufficient time for students to respond in an exploratory way in the examinations was not in the spirit of the new specifications – this was a concern in all subjects with the exception of geography. In spring/summer 2011 teachers of English literature, French, history and geography thought that the new assessment regime was encouraging independent thinking and in-depth learning. Teachers of physics and psychology, however, remained doubtful. They continued to feel that the assessment regime failed to achieve the aim of stretching and challenging students. In the final round of fieldwork in autumn 2011, heads of KS5, although recognising the opportunity for the development of subject knowledge and skills, were not unanimous in attributing this to the new specifications. Half of the heads of KS5 were not convinced that the specifications had made any difference to the depth of subject knowledge or skills developed.

As was the case when discussing issues of stretch and challenge, centres offered a variety of sometimes-conflicting views about the impact of the specifications on the teaching and learning of physics. It is, therefore, difficult to put forward firm conclusions with any degree of confidence. In terms of syllabus content, there was a more or less even balance between those that felt that there was now a greater amount of subject matter, those that felt there was somewhat less and those that felt it had remained about the same. Similarly, among those that detected change – either up or down – there was little agreement on whether this was a good or a bad thing. On the one hand, the introduction of what were regarded as more modern and more interesting topics was welcome; on the other hand, there was a sense that the amount of content made it a struggle to cover it all, particularly in the case of A2. Against this, there were regrets that students were no longer required to have some knowledge at their fingertips, for example formulae or unit quantities. But the potential advantages of such changes were that more time could be given to focusing on key concepts or to the application of knowledge.

There was wider agreement about the importance of developing skills such as the ability to analyse data or to apply knowledge in carrying out investigations or solving problems. A majority of the centres felt that the new specifications had placed greater emphasis on this and that it was to be welcomed – not least because it helped students to become more engaged with the subject. Of those that disagreed, the reasons commonly given were that students might struggle to apply subject knowledge, especially if the context was unfamiliar, or, less commonly but perhaps more surprisingly, that teachers used to ‘traditional’ ways might themselves struggle.

3.3.4 The development of higher-level generic skills

During the visits in 2010, the ability to engage in independent and investigative learning, and the capacity to think holistically, were commonly cited as examples of higher-level generic skills. These skills were seen as valuable not only in a subject-specific context, but also in enabling entry into higher education. Opportunities for their development in the specifications were felt to be provided by:

- greater flexibility in choosing course content
- more-engaging choices of topic or areas of thematic study
- greater emphasis on the application of knowledge and skills
- the synoptic units.

The follow-up visits to the case-study centres in 2011 found general agreement that the new English literature and geography specifications and assessments encouraged the development of synoptic skills; a minority of centres considered this to be the case for history, too. However, about half of the respondents who were asked about English literature and geography thought that it was too ambitious to try to develop these skills with this age group and that they would be more suited for first-year undergraduate study.

Teachers from three subjects – English literature, geography and psychology – expressed the belief that the changes to the specifications had led to a greater focus on the application of skills and knowledge. Although this was a challenge, these teachers felt that it was a desirable move, as it provided a better preparation for study at university.

During the most recent round of fieldwork, the great majority of the heads of KS5 who were interviewed thought that the new specifications offered opportunity and promoted higher-level thinking skills – such as the ability to analyse, conceptualise – and synoptic learning. One centre mentioned in particular that the reduction in the number of examinations meant there was greater coherence in the subject as a whole rather than a ‘pigeon-holed approach’.

In the case of physics, differing views were expressed about where the opportunities lie within the new specifications for developing higher-order thinking skills and about the extent to which the specifications advanced their development. Arguably, an essential feature of such skills is the ability to transfer knowledge and understanding gained in one context to apply it in another, unfamiliar, one. A further aspect of higher-order thinking is the capacity to manage and establish links within a large corpus of knowledge rather than to be able to work only with smaller units. There was broad agreement that the new specifications went some way towards requiring the development of both of these features, at least in syllabus terms, but there was little or no agreement among centres either on whether this was reflected in

assessment practice (in the examination questions and practical work) or on whether the requirements were more or less demanding than previously, or on whether more depended on the approach of the teacher.

3.4 Student engagement and participation

3.4.1 Student engagement

The majority of the heads of KS5 reported in autumn 2011 that there was generally greater engagement in the new A levels. Overall they considered this was less to do with specific content and more to do either with the development of skills or with the structure of the course. One centre cited the synoptic nature of A levels as a positive move. Only two of the centres thought that there had been no increase in the level of engagement since the new specifications.

With physics, although there was some reference to the engaging qualities of particular content (circular motion, including simple harmonic motion resonance, was one example cited), more weight was given to the opportunities students had within coursework and the optional topic in A2 to study matters of their own choosing. There was one centre, however, which suggested that the new A* award was proving to be motivating for students, and that some students had referred explicitly to wanting to achieve it.

3.4.2 Participation

3.4.3 Changes in participation

Early indications within the case-study centres in summer 2010 suggested there had been some fluctuation in participation both in terms of numbers and student profile (for example, according to prior attainment or as a result of including students who may have previously followed a vocational course). It was difficult to tell, however, whether this was directly linked to changes in specification or to other school-based reasons. A minority of case-study centres had already raised the criteria required for entry to A level psychology and English literature courses, as they had found that students with less than a B grade at GCSE struggled with the transition to AS. In a number of centres, students were taking an increased number of AS courses in order to gain a broader education – and delaying specialising further until after the AS results.

For all new specification A level subjects entries in 2011, entries have decreased very slightly since 2010 (down from 703,000 to 696,000 – see Figure 5), following four previous

years of increasing numbers since 2007.[45] Figures 6 to 11 show that there are variations from subject to subject, with entries increasing for physics, for example.

The proportion of male and female entries has remained constant across the period from 1996 to 2011, with females making up around 55% of all entries.

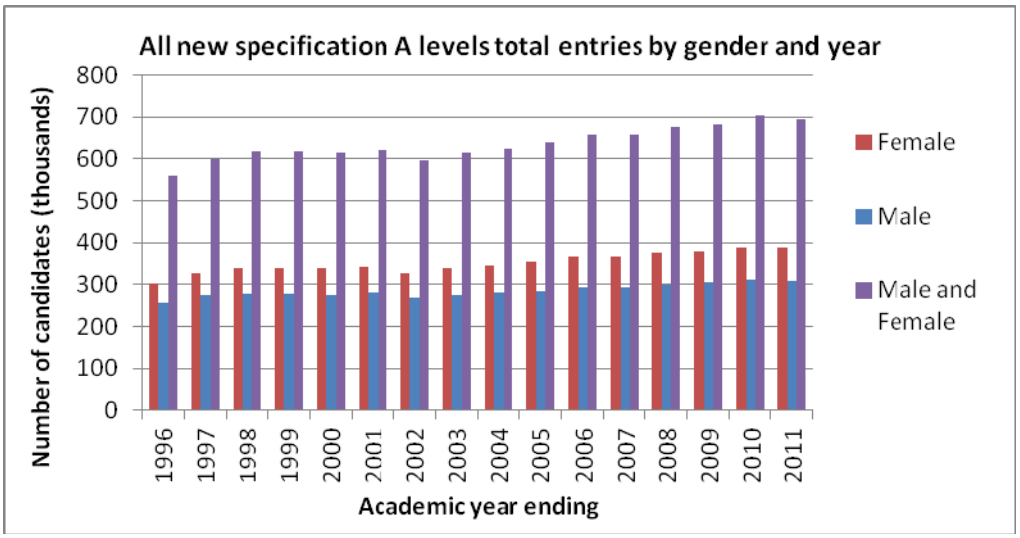


Figure 5: Entries for all new specification A level subjects, 1996–2011 (SFR)

3.4.3.1 French

In French, entries at A level have decreased a little to the lowest level since the SFR records began in 1996. The proportion of male and female entries has remained constant throughout, at around 69% female (Figure 6).

[45] Since students of school age (i.e. 16–18) are of most interest in this report, the numbers are based on the SFR data.

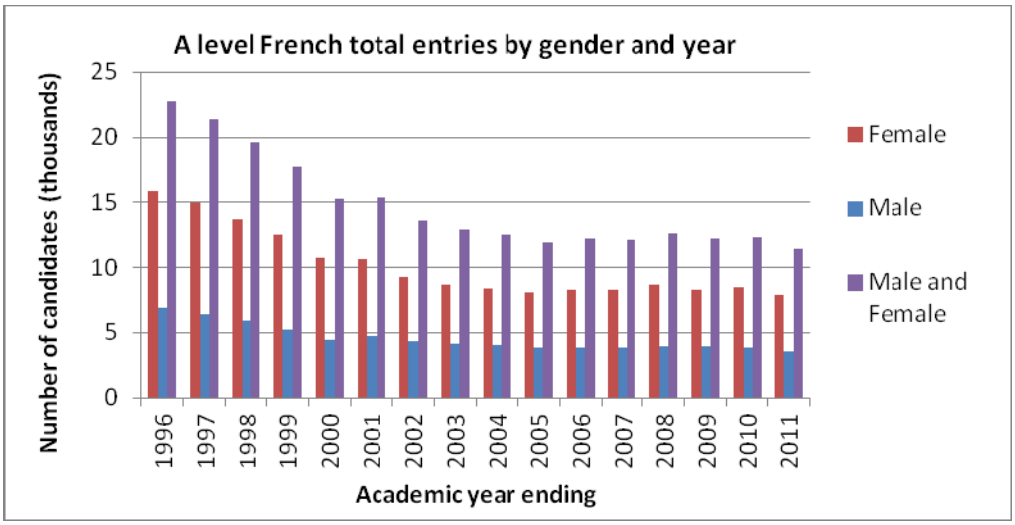


Figure 6: A level French entries, 1996–2011 (SFR)

3.4.3.2 Geography

In geography, entries at A level have decreased a little, despite increasing in the previous four years. The proportion of male and female entries has remained largely constant throughout the period shown, at around 45% female (Figure 7).

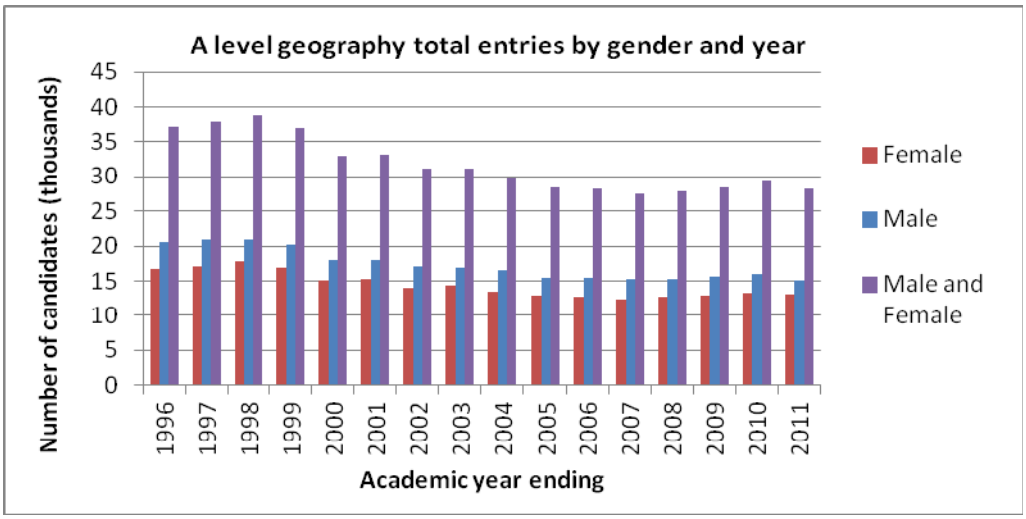


Figure 7: A level geography entries, 1996–2011 (SFR)

3.4.3.3 History

In history, entries at A level have continued to increase. The proportion of male and female entries has varied somewhat over the years, with a higher proportion of female candidates before the Curriculum 2000 reforms (an average of 55% from 1996 to 2001), falling to an average of 50% from 2002 to 2011 (Figure 8).

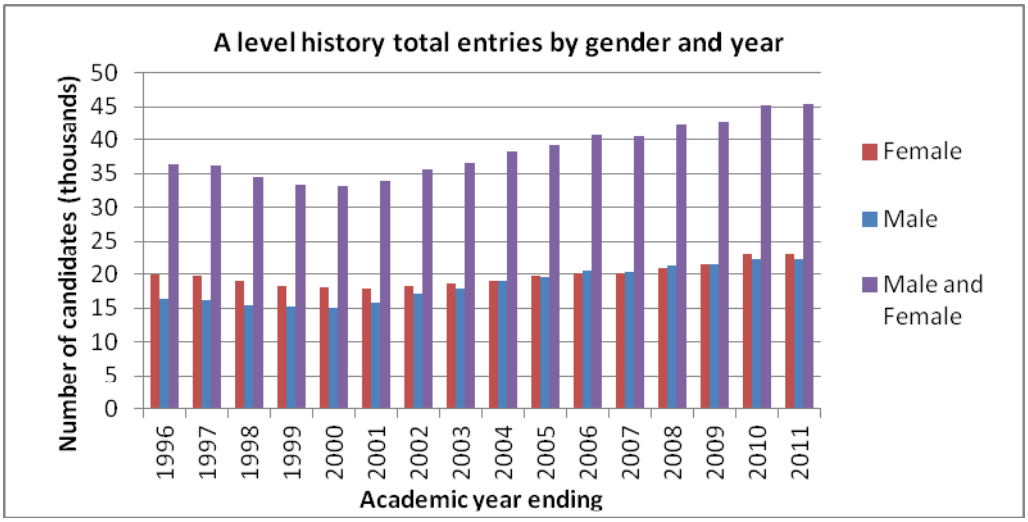


Figure 8: A level history entries, 1996–2011 (SFR)

3.4.3.4 Physics

In physics, entries at A level have continued to increase, with entries now at the highest level since 1999 (29,216). The proportion of male and female entries has remained constant throughout, at around 22% female (Figure 9).

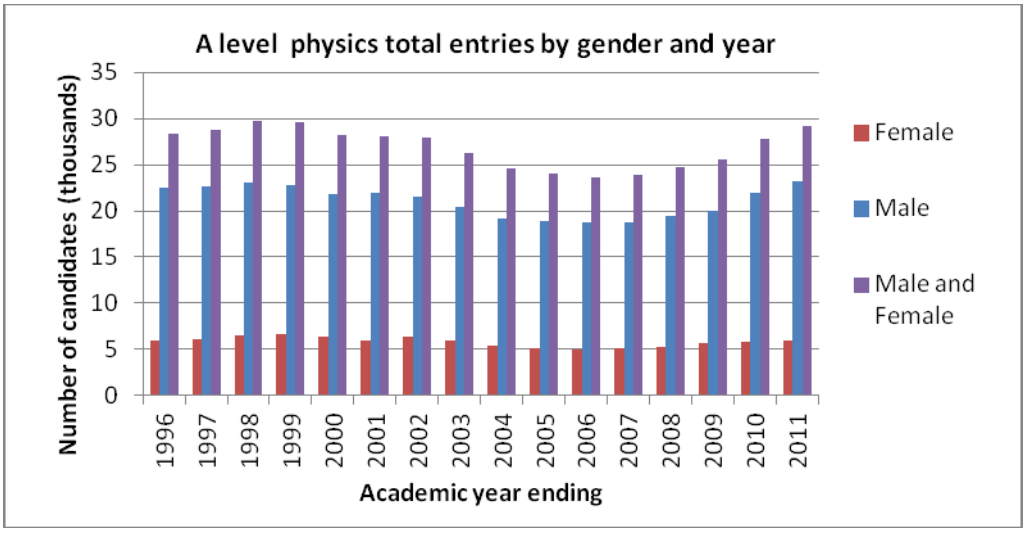


Figure 9: A level physics entries, 1996–2011 (SFR)

3.4.3.5 Psychology

In psychology, entries at A level have plateaued after a generally increasing trend since 1996. Entries have remained similar in 2011 to that of 2010. The proportion of male and female entries has remained constant throughout, at around 75% female (Figure 10).

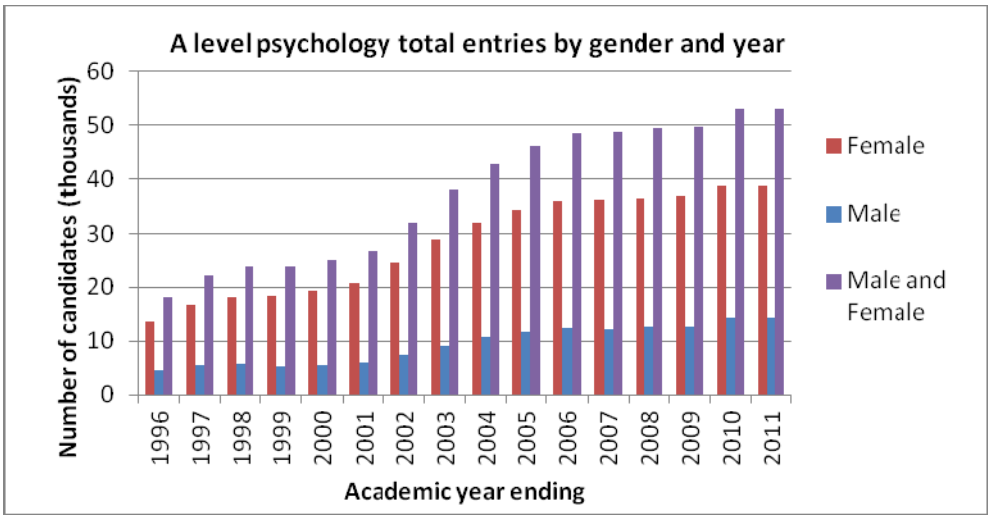


Figure 10: A–level psychology entries, 1996–2011 (SFR)

3.4.3.6 English

Data is not available for English literature individually, so the charts displayed here represent A level English subjects as a whole, including results for English language, English literature and English language and literature. Around 55% of the UK's English A level results in 2010 were from English literature (based on data from JCQ). Assuming similar re-sit patterns and participation trends between different years and cohorts, it is likely that around 50–60% of the English entries reported here are entries in English literature. At A level, entries have fallen. This decrease is mainly due to a decrease in the number of male entries between 2010 and 2011; female entries have remained constant (Figure 11).

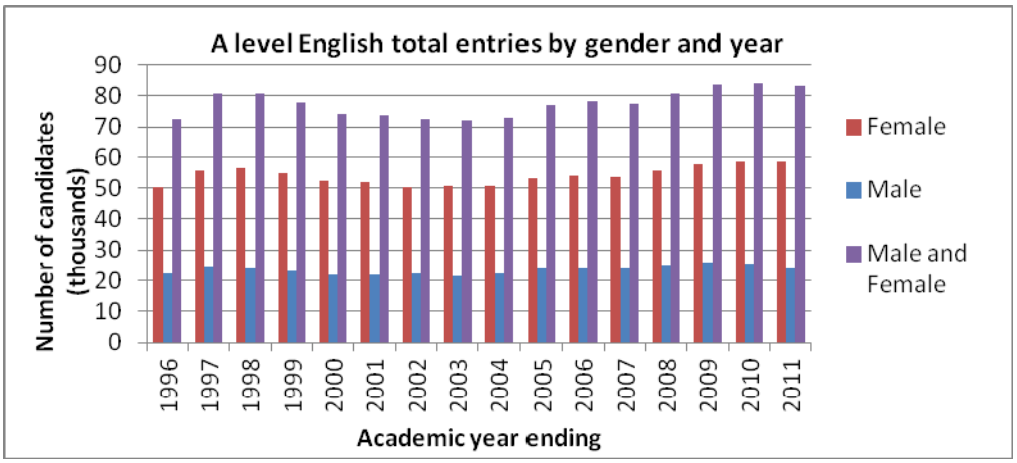


Figure 11: A level English entries, 1996–2011 (SFR)

3.4.4 Re-sitting

A level students interviewed welcomed the opportunity for re-sits, with the majority believing that re-sits should be allowed in any circumstance and only a minority suggesting that re-sits should be allowed in only exceptional circumstances.

The statistical data, however, indicated that opportunities for re-sitting were likely to be less beneficial than students suggested. Although the 2010 ABDA data lacked the wealth of information about re-sits of the 2008 data, some insight was gained by examining the change in uniform mark scale (UMS) scores between re-sits of unit 1 at series 1 (i.e. the first attempt) and series 2 (the first re-sit).[46] Re-sits of unit 1 increased from 7.1% of the candidature in 2008 to 9.6% in 2010.[47] If re-sitting leads to an improvement in grades, and with more re-sitting in 2010, it might be expected that UMS scores (and grades) on average would have increased more between the first sitting and the re-sit in 2010 than in 2008. This was not found to be the case. In both years the effect of re-sitting was small (an average gain across the re-sitters of 7 UMS points for the unit), but the gain was smaller in 2010: an average of 5 UMS points for the four subjects where an improvement was seen – English, geography, psychology and physics. Only for geography was the average re-sit improvement higher in 2010 (5.7 UMS points) than in 2008 (4.1 UMS points). The 58 re-sitters of unit 1 in French in 2010 actually performed less well on average than in their first sitting. Unfortunately, analysis of trends across a wider range of years is not possible, as re-sit data is only available for 2008 and 2010.

Stakeholders expressed two particular concerns, one more general than the other. The first was the undesirable impact that retake opportunities were considered to be having on learning. Retakes, it was argued, encourage centres to spend a lot of time preparing students to re-sit examinations in January with the aim of achieving a better grade – this may indeed be the outcome, but, if there were no opportunities for re-sitting, more time would be available for teaching and learning and there would be no real change to the differentiation between students' achievements.

The more specific concern was where courses involve a practical element. Rather than being examined in 'real time', as previously, under the new arrangements the practical examination has no fixed length and can be repeated several times until students reach their target grade. It was felt that this led to a situation where students were not being given a true test of their ability.

3.5 Progression

In 2010, for half the subjects, case-study centres reported that there had been fewer students progressing from AS to A2 since the introduction of the new specifications. The decrease had been most dramatic in psychology, where the introduction of a science focus

[46] A level history is not included in the ABDA data.

[47] The years here refer to the year of completion of the A level.

had led to only about half the student cohort continuing after AS. French and geography had also seen a reduction in the number of students progressing from AS to A2. There had been no significant changes in the other three subjects.[48] English literature case-study centres reported no particular effect of the changes on the rates of students progressing from AS to A2, noting that these rates continued to be as healthy as ever.

During the case-study visits in 2011, only the French AS was seen as more demanding than before and leading to improved preparation for A2: the level of difficulty of the reading texts at AS was reportedly higher than those at A2 French. Two centres also suggested that the standard expected of students in GCSE French ought to be raised, to prepare students better for the new AS French.

For English literature, history and geography, those who responded thought that the AS was preparing students less well for A2 than before. The reasons given were lack of subject-specific skills development (English literature), lack of variety and too much crammed in (history), and lack of emphasis on essay-writing skills (geography). Progression from AS to A2 was considered by many case-study centres to be a larger jump than in the previous specifications. During the case-study visits in 2011, for example, for English literature, history and geography those who responded thought that the AS was preparing students less well for A2 than before.

The statistical data discussed below indicates a drop in the proportion of students progressing to A2 after AS. The emerging pattern seen is broadly in line with what the case-study centres suggested about the extent to which AS was good preparation for A2 study.

As noted in Section 2 (Sources of evidence), SFR AS data is incorrect as a combined result of the way in which some students cash in their AS attainment and the way in which qualification discounting is operated in the SFR and NPD datasets. The implications of this are that:

- the number of candidates taking AS level subjects in any given year is under-reported in the SFR
- the implied progression of AS candidates to A2 is over-estimated, typically by about 20%.

The complexity of A/AS level sitting patterns means that the progression figures presented below remain an estimate of actual progression. It can be seen that for specifications under consideration (except for French), progression rates have fallen by around 3%. However, it

[48] History was not included in 2010.

should be noted that progression for mathematics, which did not undergo a specification change in 2010, has fallen too. Whether these changes relate to the specification change or other factors is not known.

Table 3: Proportion of candidates who took AS and then went on to complete A level in 2009 and 2010 (%) (NPD data)

	Year of A2 completion	
A level	**2009**	**2010**
Physics	65.0	62.0
Mathematics	71.9	68.0
Geography	72.0	71.1
Psychology	63.5	61.5
English literature	75.2	73.7
French	63.5	63.3
All new-specification subjects	64.2	62.8

Note that:

1. Checks for non-existent but expected AS results have been undertaken only for the year preceding the A2 result (not for previous years).
2. The 'All new-specification' A levels includes only those A levels where 2009 and 2010 completions are comparable (so excludes A levels that are newly introduced qualifications in 2010 or qualifications phased out in 2009).
3. Across the 74 new-specification A levels considered, 54 (73%) have lower proportions completing in 2010 (the first year of the new specification) than in 2009.

Teachers felt that A2 English literature, geography, psychology, history and French were more challenging, and better preparation for higher education – for example, in terms of encouraging independent learning. There were some exceptions in terms of specific skills development for a subject – for example, some geography teachers felt that there was not enough emphasis on the research skills required for university study. The A* grade at A level is perceived by both teachers and students as recognising high-attaining students. Some universities now require an A* pass for certain courses.

3.6 The impact of change on centre behaviour

3.6.1 Choice of specification

Although not always stated explicitly, the three main reasons for a centre's choice of one specification rather than another were: that it best met the needs and aspirations of their students; that it was the one that was most consistent with how they viewed the nature of the

subject; familiarity – often long-standing – with a particular awarding organisation. These reasons were uppermost in the responses across the range of subjects, but there were, within them, a number of more-specific considerations.

3.6.1.1 The needs and aspirations of students

Three of the centres expressed concern that the specification should be accessible and student friendly. The characteristics of accessibility and student friendliness were seen in terms of:

- course content – for example in geography, where students undertook a local area study
- course structure – for instance in modern foreign languages, where incremental assessment was seen as being fairer on students; in coursework, which, in history, offered more variety or, in English, allowed students an element of independent choice; and where a centre welcomed the existence of the synoptic unit in English
- the perceived interest of the course for students, in relation to modern foreign languages and history
- the course's appropriateness for students with a range of attainment in a given subject (the reference here was to physics).

3.6.1.2 Consistency with the nature of the subject

In terms of consistency with the subject's nature, the particular features that were referred to were:

- studying a particular period rather than random texts in English
- in modern foreign languages, the linguistics-based nature of the course and the choice of topics in French
- the requirements of the practical component in physics
- a more issues-based specification in geography.

3.6.1.3 Familiarity with the awarding organisation

Case-study centres cited familiarity as a reason to remain with an awarding organisation. Among other advantages it was felt that this would make administration simpler. Allied to familiarity with the awarding organisation were considerations about teachers' expertise and experience (for example, the specification allowed teachers to follow their interests) or that it was the better option because it related more readily to teachers' specialisms.

3.6.2 Management and resource implications for centres

There appeared to have been little impact on management time as a result of the changes to the A level specifications. Where centres reported differences, they tended to focus on the impact of coursework on workload rather than the management of A levels. Any increase in the resource burden was also seen as adding to the workload of subject teachers. One centre, for example, felt that – as a result of cuts in funding and the reduction in examination officer time – teaching staff were taking on more of the burden of examination administration, which added to their workload. In the online centre survey, managing and structuring the whole scheme of work for the A level course was seen as the single most significant change that respondents reported seeing since the introduction of the new A level specifications.

A number of respondents to the online survey had reported implementing greater use of IT and/or online resources when moving to the new specifications. In particular, respondents in modern foreign languages spoke about the greater use of MP3 players and computers in learning and in examinations.

4 GCSE findings

4.1 Attainment

4.1.1 Grade outcomes for new-specification GCSEs (2009 first teaching)

Overall, mean grade score has decreased from 5.27 in 2010 to 5.23 in 2011. To recap, for GCSE grades, an A* grade is weighted as 8, A as 7, B as 6, C as 5, D as 4, E as 3, F as 2, G as 1, and all other grades (i.e. U and X) are given zero weighting. Thus, a mean grade score of 5.23 corresponds to a low C grade. Mean grade score has decreased by a greater amount for males than it has for females: for males, mean grade score fell from 5.02 in 2010 to 4.96 in 2011; for females, mean grade score decreased only slightly from 5.49 in 2010 to 5.48 in 2010. This decrease is noteworthy as it follows a period of steadily rising grades since 2004.

The difference in attainment between males and females has increased, with females in 2011 having a mean grade score of 0.52 points more than males. This confirms that the trends observed for the four new-specification GCSE qualifications reviewed below are indicative of the outcomes for new-specification GCSEs as a whole. What is interesting for this evaluation is that, although outcomes at GCSE for male entries are consistently weaker than for females, this gap has widened significantly with the introduction of the new specifications (rising from a difference of 0.46 of a grade in 2010 to 0.52 of a grade in 2011).

In addition, the outcomes for both sexes have dipped in 2011, following a period of steadily improving grades since 2004.

The charts below (Figures 12–14) show the mean grade scores for all new-specification GCSE entries in 2011 (and preceding years) based on JCQ entries data. All GCSEs are included except English, mathematics, science (separate sciences and combined awards) and ICT.

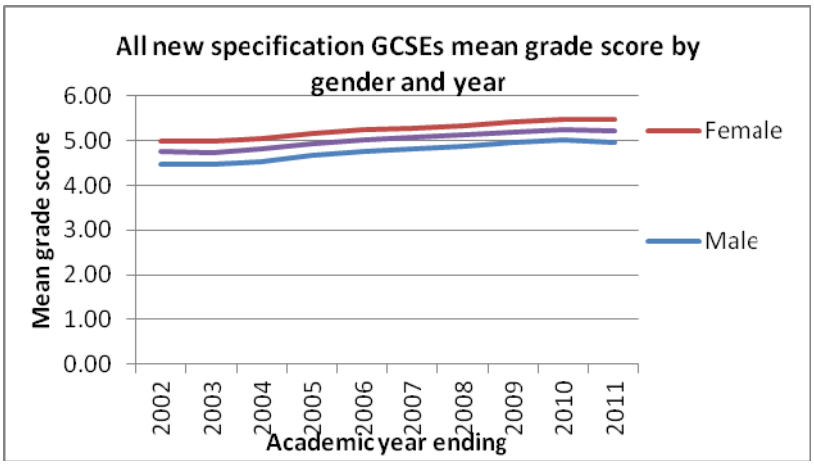


Figure 12: GCSE all new specifications mean grade score, 2002–11 (JCQ)

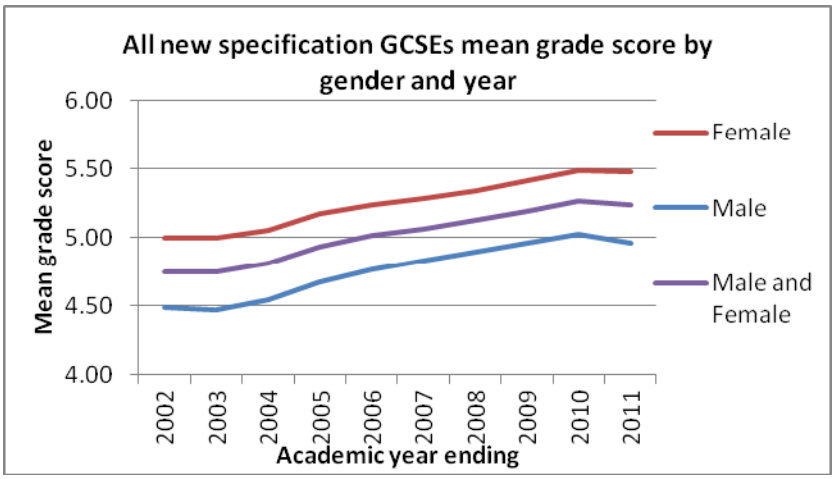


Figure 13: GCSE all new specifications mean grade score, 2002–11 (JCQ) (zoomed y axis – grade score – to show change more clearly)

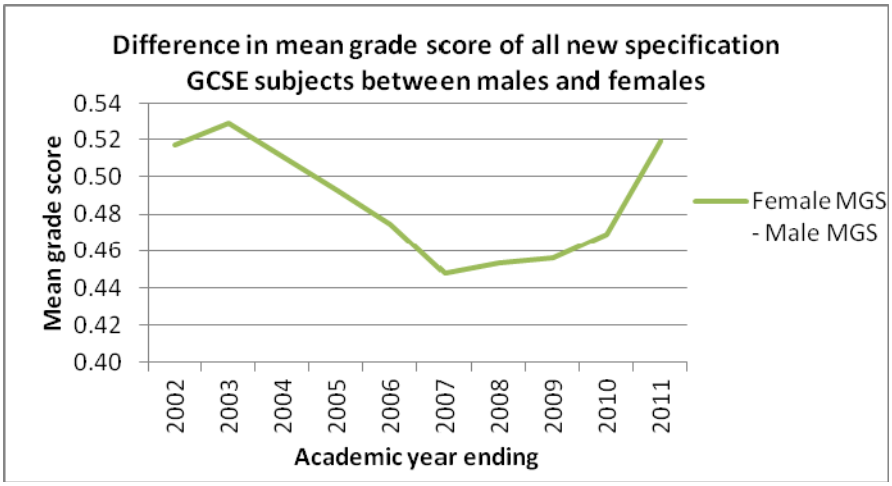


Figure 14: Difference in GCSE all new specifications mean grade score, by gender, 2002–11 (JCQ)

4.1.1.1 French

Male GCSE results in French have on average worsened, while female GCSE results have continued to improve at the same rate seen in previous years. The overall mean grade score has continued to increase (very slightly greater at 5.39 in 2011 compared with 5.38 in 2010), although the rate of increase has fallen – a plateauing effect created by the combination of females' attainment increasing as predicted by previous years' results, and males' attainment falling off a little (Figure 15).

Since 2003 the gender gap has been gradually closing, but the 2011 results have re-opened the gap to a level not seen since 2006, with the difference in mean grade score increasing from 0.32 in 2010 to 0.43 in 2011. Note that each point in mean grade score corresponds to one grade, so here a difference of 0.43 corresponds to close to one-half a grade difference in the average performance of females and males (Figure 16).

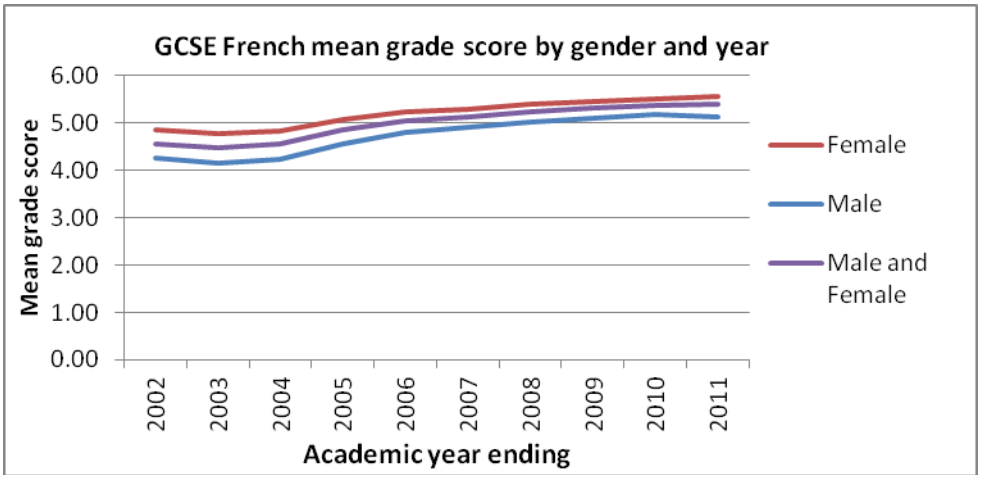


Figure 15: GCSE French mean grade score, 2002–11 (JCQ)

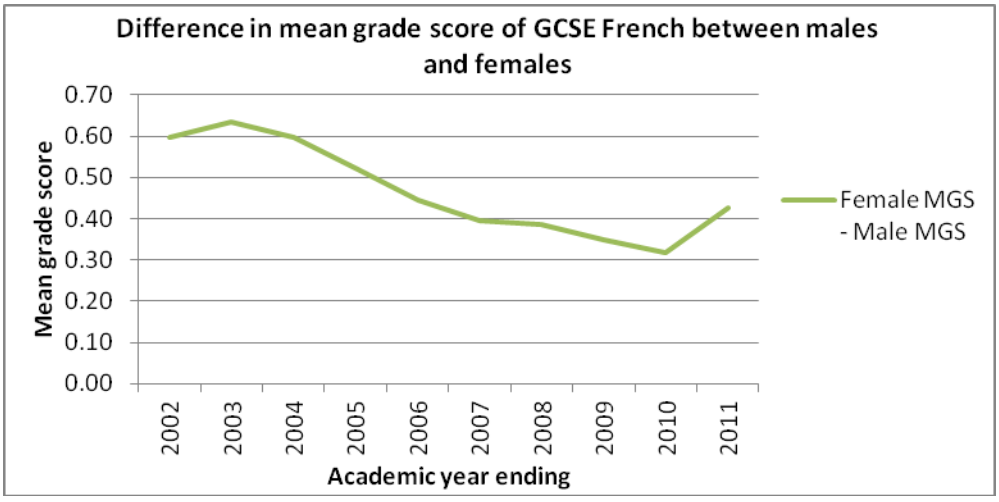


Figure 16: Difference in GCSE French mean grade scores, by gender, 2002–11 (JCQ)

4.1.1.2 Geography

Males' attainment in geography has on average worsened, while females' attainment has continued to improve at a similar rate to that seen in previous years. The overall mean grade score has continued to increase (very slightly, from 5.30 in 2010 to 5.31 in 2011), but the rate at which this value has increased is lower than in previous years – a plateauing effect created by the combination of females' attainment increasing as predicted by previous years' results, and males' attainment falling off a little (Figure 17).

The difference in average performance between males and females has increased considerably since 2010. The gender gap, which was already on the increase since 2009, increased by almost a third in just one year: the mean grade score of females was 0.35 higher than that of males in 2010, whereas this difference now stands at 0.46 (Figure 18).

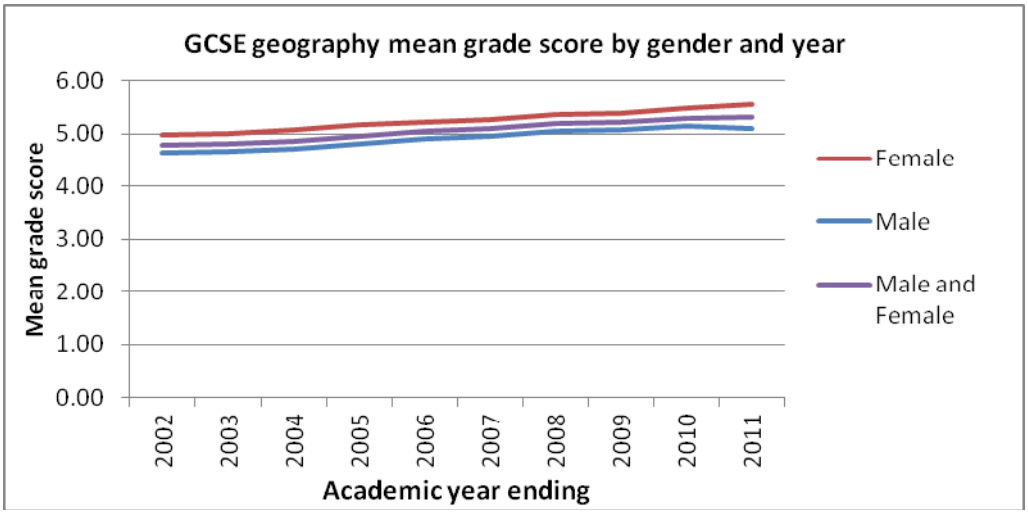


Figure 17: GCSE geography mean grade score, 2002–11 (JCQ)

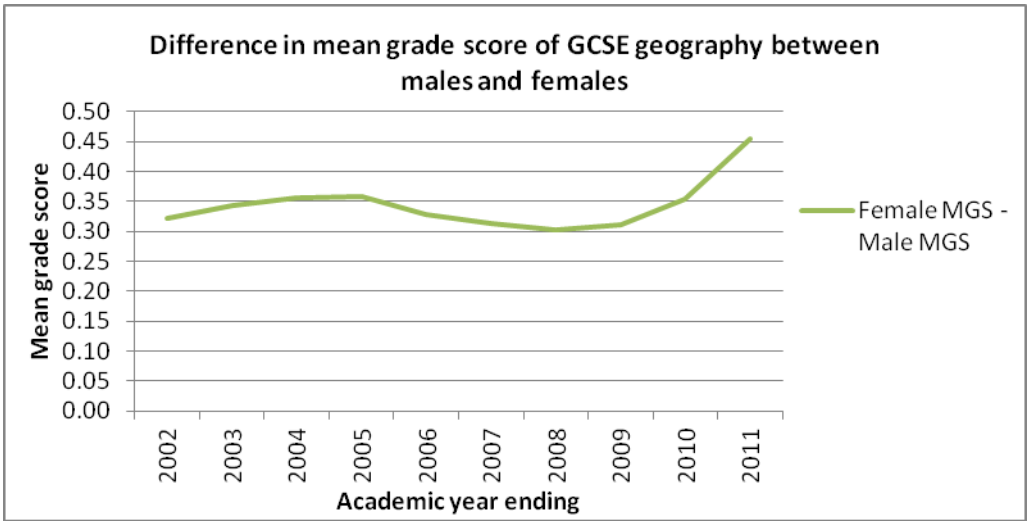


Figure 18: Difference in GCSE geography mean grade scores, by gender, 2002–11 (JCQ)

4.1.1.3 History

Despite following a consistent upward trend since 2002, neither male nor female attainment in history increased between 2010 and 2011. While females' average attainment remained constant, with a mean grade score of 5.53 in both 2010 and 2011, males' average attainment has decreased from 5.14 in 2011 to 5.11 in 2010 (Figure 19).

The difference in mean grade score between males and females has remained reasonably constant, although it was at its highest level (at 0.42 of a grade) in 2011 (Figure 20).

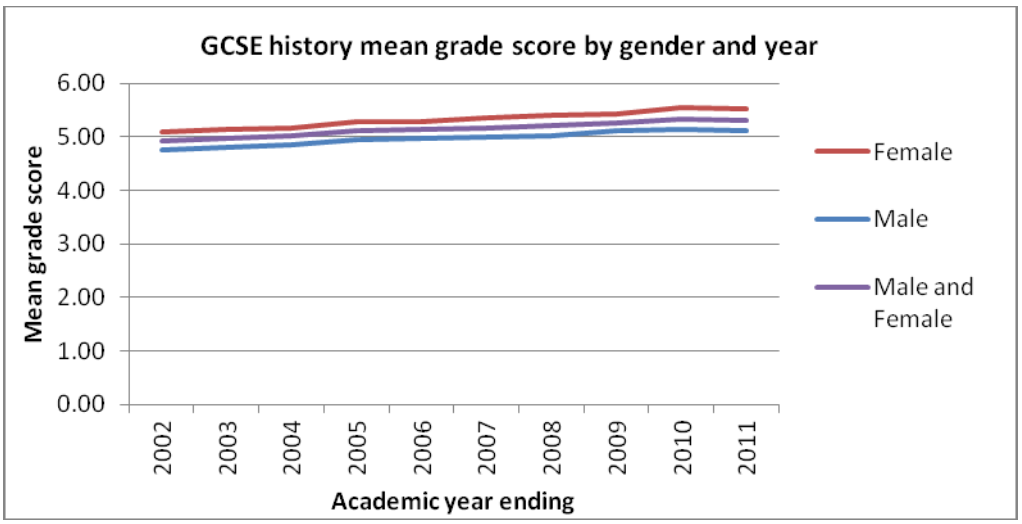


Figure 19: GCSE history mean grade score, 2002–11 (JCQ)

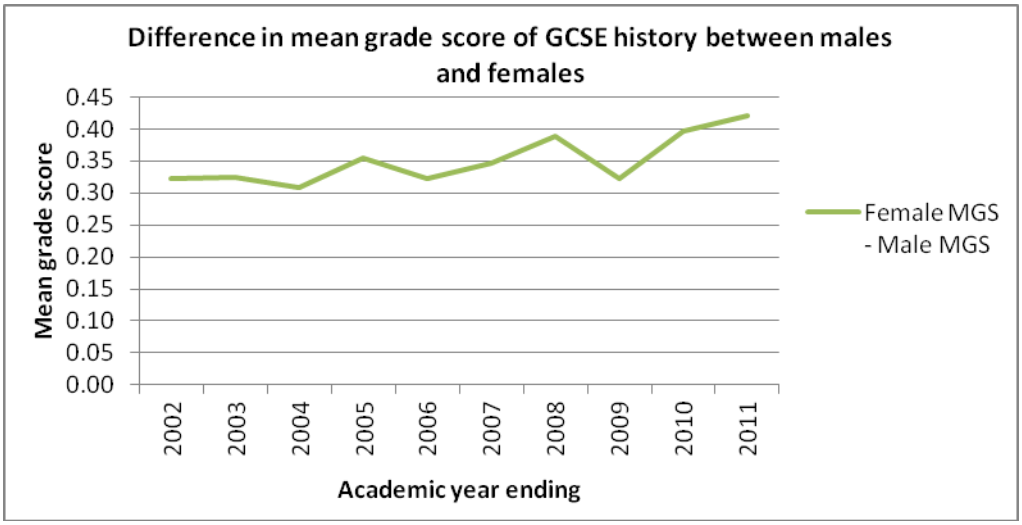


Figure 20: Difference in GCSE history mean grade scores, by gender, 2002–11 (JCQ)

4.1.1.4 Spanish

Females' attainment in Spanish decreased from 5.77 in 2010 to 5.75 in 2011 (a drop of 0.3%) and males' attainment fell from 5.42 in 2010 to 5.27 in 2011 (a drop of 2.8% – see Figure 21).

Despite continuing to narrow since 2002, the gender gap has increased between 2010 and 2011. The difference in mean grade scores has increased from 0.36 in 2010 to 0.49 in 2011 – a difference not seen since before 2005 (Figure 22).

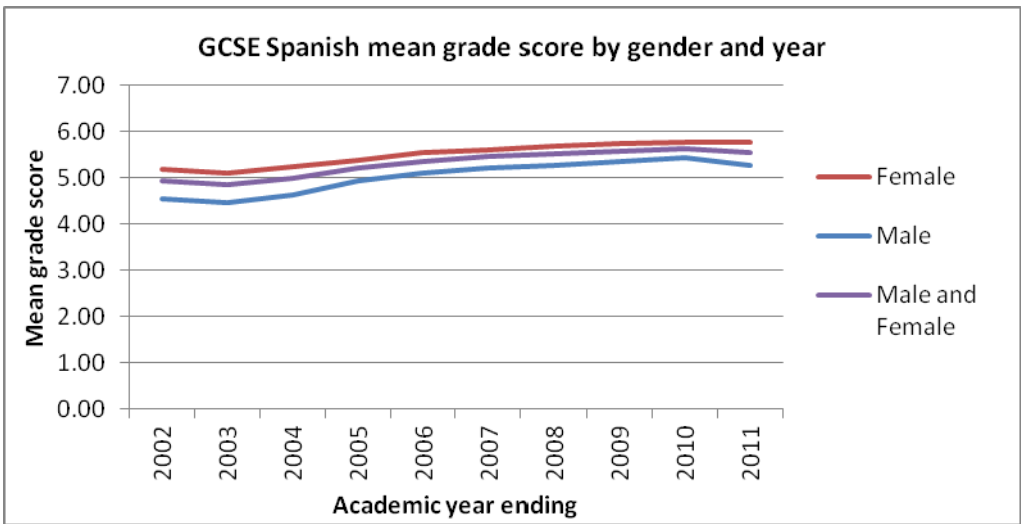


Figure 21: GCSE Spanish mean grade score, 2002–11 (JCQ)

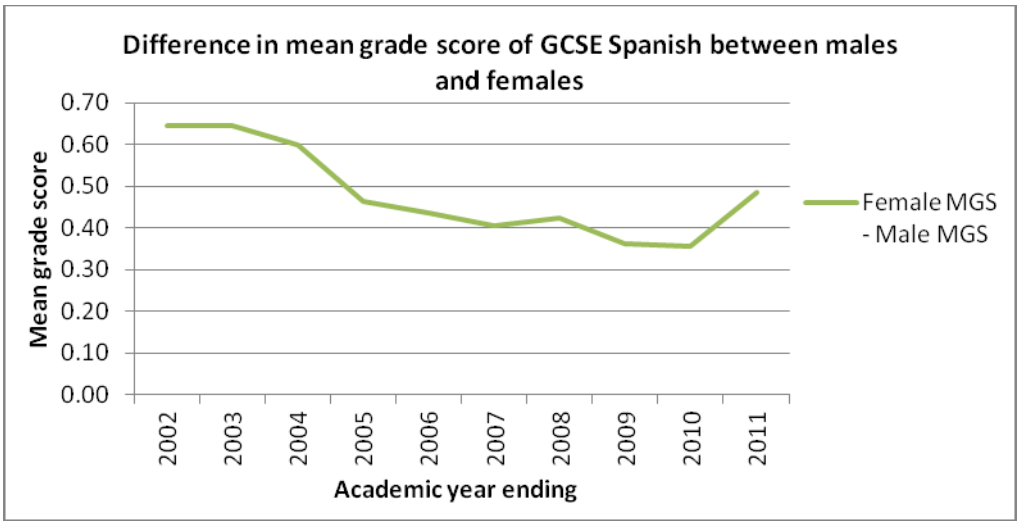


Figure 22: Difference in GCSE Spanish mean grade scores, by gender, 2002–11 (JCQ)

Attainment data for English/English language and mathematics has not been included here as there is no awarding for the new-specification GCSEs until 2012. Data for the legacy qualifications can be found in Appendix 3.2.

Summary points

Two important caveats need to be placed on these findings:

- data is for entries rather than candidates, which means that the effects of re-sitting may cloud information about overall outcomes
- no data is available yet for candidate characteristics.

Nevertheless, on the basis that there is no reason to suspect either major changes in re-sitting behaviour since 2010, or any differences between the 2011 year 11 cohort and the 2010 cohort in terms of prior attainment or other important characteristics, it is possible to draw two tentative conclusions:

- The overall grade distributions for males have got worse (although the change is relatively small) for all four new-specification GCSE subjects under consideration: French, geography, history, Spanish. These changes have not been identified in English and mathematics, for which specifications had not changed in summer 2010. For the four new-specification subjects, the gap between average male grade and average female grade (females performing consistently better over time in each subject) widened in 2011.
- The overall grade distributions have plateaued for all candidates in GCSE French and geography, and decreased a little for GCSE history and Spanish.

This allows the following provisional conclusions to be drawn:

- The new GCSE specifications are proving a little more challenging for males than females, even taking into account attainment differences in previous years of GCSE.
- There is a similar plateauing effect as was seen for A level in 2010, suggesting that Ofqual's revised guidance to awarding organisations has had the effect of maintaining similar overall grade distributions in a situation where grades might have been expected to get worse, based on past experience at times of specification change.

It should be noted that more-detailed analysis of a wider range of new specification qualifications is needed before this last conclusion may be confirmed, coupled with a consideration of candidates' prior attainment for the years before and after the specification change.

4.1.2 Centre perceptions of attainment

In spring 2011, much of the dialogue on attainment at GCSE was based on teachers' predictions for controlled assessment and examinations, as opposed to actual results. The overall feel was that attainment would remain at a similar level as before. Where centres had had results for controlled assessments, these varied across and within subjects, although several centres reported a downward trend on their predicted grades for GCSE geography.

In the online centre survey, of those that had received results for controlled assessment, 12 reported that the results were as expected, 4 that they were below expectations, a further 4 that they were significantly below expectations, and 2 that the results were above expectations. The subjects for which results were reported as being significantly below expectations were GCSE French and Spanish. The response size here was so small, however, that no reliable conclusions could be drawn.

There were seven further interviews with heads of KS4 in autumn 2011, following the first awards for the new specifications (not including English and mathematics). Five of these reported that the 2011 grades were not in line with their predictions. For three centres grades were worse; two centres reported better grades overall, although one of the two reported some underlying problems that had led to a specific fall in geography and history results. Four of the five had seen some difference in grades for a particular group or cohort of students: three in relation to the lower-attaining groups, including SEN students and the least motivated students, and one in their middle-ability group. They suggested that the lower-ability students had done worse and the middle-ability students better than before.

4.2 The impact of changes on teaching and learning

4.2.1 The development of subject specific skills and knowledge

In the online centre survey conducted in spring 2011, of the 51 respondents nearly half (23) did not feel that students were necessarily experiencing a deeper, more-lasting understanding of the subject as a result of the changes to the GCSE specifications, and 17 felt that there was no change. Only 11 respondents (English, French, geography and mathematics teachers) felt that students were gaining in this way. There was no subject-level agreement about the impact – teachers of the same subject expressed differing views.

These latter views – that the specifications had had no impact or had had a negative effect – were also reflected in the responses of the seven heads of KS4 interviewed in autumn 2011.

The following general points emerged about specific subjects.

- With both geography and history, the general view was that the new specifications did not encourage the development of subject-specific skills. In geography, the reason given was that covering all physical geography topics in one year and human geography topics in another year failed to engage students and did not help them to develop their data response skills – something to which the physical modules do not lend themselves well. In history, the reasons cited were that the course was too easy overall and that the unitised model of assessment did not give students enough time to develop their skills.
- A minority of modern foreign language (MFL) teachers and heads of department thought that the new speaking test in GCSE French and Spanish enhanced the development of subject-specific skills by guarding against students simply regurgitating pre-learned material. It was also noted, however, that the test expected a degree of spontaneity beyond GCSE level and that the reduced emphasis on grammar skills and on 'passive skills' (listening and reading) in the assessments disadvantaged lower-attaining students.
- The five heads of English interviewed in autumn 2011 thought that changes in the development of subject-specific knowledge and skills had been only minor.[49] Examples where teachers had made changes to their teaching as a result of changes to subject-specific skills and knowledge included the introduction of the

[49] It should be noted that the heads of English did not separate between the English language and literature specifications, which they consider a dual award, not two separate specs. Although the focus for the evaluation is primarily English and English language, separating English language and English literature appears to be against school practice and against the way in which heads of English perceive them.

analysis of spoken language and the unseen question on poetry. However, it was also noted by some that the English literature GCSE was too content heavy and that this resulted in students having insufficient opportunity to develop their subject-specific skills.

- Overall, the four heads of mathematics also interviewed in autumn 2011 thought that the subject knowledge or what students need to know mathematically for the new specification was very similar to what it had been before. There was also consensus that the most substantial change was the introduction of more functional or applied mathematics, resulting in a significant change to the subject-specific and generic skills now required from students, and thinking and problem-solving skills in particular. However, acquiring these skills was deemed to be very difficult for students with weak literacy skills. It was noted that lower-ability students would benefit from more-basic mathematical skills over needing to tackle functional problems.

4.2.2 Opportunities for higher-level skills development

The broad consensus that emerged with all subjects was that the new specifications were aiming to promote and provide opportunities for the development of higher-level skills, but that there were some inconsistencies and some gaps between aims and practice. The heads of KS4 who were interviewed, for example, thought that there was a dislocation between the specifications and the forms of assessment. The specifications promoted the development of higher-level thinking skills as much as or more than the previous specifications, particularly with their emphasis on the skills associated with independent learning and thinking – but this was not reflected in the examination papers and criteria. Dislocation was not a concern shared by other respondents, but they did raise further issues. In the case of mathematics, it was thought that the style of questioning and the assessment criteria for the functional and problem-solving questions did promote independent thinking, as students needed to think through a problem and break it down into stages. The heads of mathematics departments interviewed, however, questioned why the specifications included some but not other higher-level thinking skills. There was a greater focus on the ability to analyse, to question and to solve problems, for example, but not on the development of the ability to conceptualise or of synoptic learning skills. The majority of heads of department also had concerns that the development of higher-level thinking skills in the new specification worked well for higher-tier students but not for foundation-tier students, and they had further concerns over the level of literacy that students needed to tackle functional and problem-solving questions.

With specific reference to geography, three case-study centres expressed the view that the unitised approach to assessment in GCSE does not encourage synoptic learning, which is

an important element in developing higher-order skills. It was, therefore, up to teachers to make the connections between themes for the students.

4.2.3 Depth of learning

The majority of heads of KS4 argued that deep learning and understanding (which the new specifications seek to promote) come not from increased volumes of content, but from enabling students to develop a given set of skills. This was a view shared by most subject heads. There were therefore concerns that the amount of content in many specifications restricted the opportunities for students to apply their skills and thus deepen their mastery of them. Deep learning was seen to be promoted in particular where project work was part of the specification.

Respondents were more or less equally divided in their views about whether the new specifications clearly promoted deep understanding of subjects, or only to some extent, or not at all. The great majority of heads of English thought that, although some elements of the new specifications promoted deep learning, other elements 'negated' it. The elements seen to promote deep learning were the increased focus on functionality of English, the ability to write for audience and purpose, and the spoken language unit (in which students do their own research). Conversely, factors leading to a lessening of deep learning were the facts that students were no longer required to study whole texts and that the 'over-packing' of the specifications was leading to less opportunity for teachers to promote good teaching and learning strategies in their lessons.

Similarly, the heads of mathematics departments felt that, although the new specification for the single GCSE in mathematics promoted deep learning, it did so in varying degrees. The elements cited as promoting deep learning were the introduction of functional mathematics, problem solving, and questions with less 'scaffolding' (so that students had to break questions down and decide on an approach for themselves). The majority of heads expressed the concern, however, that lower-ability students were in a position to develop neither those extra skills, nor an understanding of why one method of resolving a particular problem rather than another works, and how an answer might be found.

4.3 The role of assessment in teaching and learning

The introduction of the new specifications was seen to have increased the focus on assessment in teaching and learning. In addition to making this point generally, the focus on assessment came through in the responses to questions about specific areas of teaching and learning.

4.3.1 Impact of unitised vs. linear assessment

In spring 2011, 5 of the 14 case-study centres discussed the respective merits of unitised and linear assessment in GCSE French, history and mathematics. They were of the view that linear assessment promoted more in-depth and longer-lasting learning than unitised assessment, particularly in relation to the development of subject-specific skills. Similar views were expressed by wider stakeholders.

In the senior leaders online survey, there was a fairly even spread among the respondents between those who had opted overall for linear assessment, those who had opted for unitised assessment and those who had opted for a combination of both. Approximately half of those providing an explanation indicated that the choice of unitised or linear assessment was a pragmatic one that depends mainly on the nature of the subject.

There were varying and sometimes diverging views on the merits of unitisation expressed by the senior leaders surveyed in spring 2011. Respondents' concerns included a belief that unitisation destroys coherence within subjects and does not encourage students to see how the parts 'fit together'. There was also a view that unitised examinations (and re-sits) may improve results, but do not help with a stronger understanding of the subject – too much time is spent examining, and not enough on teaching and learning. Conversely, the supporters of unitised assessment argued that it allows students to achieve along the way – and to re-sit if necessary.

The students were in favour of the unitised approach for a number of reasons: it made learning easier (as it was presented in bite-size chunks), the content was clearer, and the material was fresher in the learner's mind when it came to the examination. Students also liked the fact that unitisation meant that there were exams throughout the year rather than a single exam at the end. They felt that this not only took away some of the stress of examinations, but also that it was easier to revise and it made all the topics link together better, as students learned a whole topic in one section before moving on.

4.3.2 Impact of controlled assessment on teaching and learning

The purpose of controlled assessment is to assess those aspects of a subject that cannot be easily assessed by external examination. These aspects include research, planning, investigation, analysis, collaborative working, and presenting ideas and arguments supported by evidence.[50] Controlled assessment is designed to encourage a more

[50] QCDA (2010) *Managing GCSE controlled assessment – a centre-wide approach.*

integrated approach to teaching, learning and assessment, and to enable teachers to confirm that students carry out the work involved.[51]

By late spring 2011, a wide range of approaches to controlled assessment had been implemented within and across centres, often depending on the subject being assessed and/or on the teacher. Awarding organisations have commonly interpreted Ofqual's regulations differently, with the result that the level of control required in the regulations for the same subject may vary across organisations. Consequently, there has been considerable variation in the amount of support that teachers believe they can give their students in the preparation stage of controlled assessment.

Concern was expressed that some teachers are using strategies, even if often unwillingly, to help students to get the best grade, and that these undermine the validity of the controlled assessments in terms of the knowledge and skills that are meant to be assessed. Examples of such strategies include running practice assessments that change little when it comes to the 'real' assessments and, in modern foreign languages, teachers setting more controlled assessments than required and choosing the best for each individual student.

Most teachers liked the opportunity that controlled assessment gave them to choose topics and contextualise the tasks for their students. However, there was no consensus across or within centres and subject areas about the extent to which controlled assessment is an appropriate form of assessment for particular skills and knowledge. In spring 2011, concerns were expressed by some teachers that, rather than promoting in-depth independent learning, controlled assessment may test only the ability to learn and regurgitate a body of content. A number of MFL staff and students, in particular, felt that controlled assessment was often just assessing students' ability to memorise and regurgitate a text.[52]

The majority of heads of English echoed this view in autumn 2011, stating that controlled assessment was having a detrimental effect on deep learning. In their view, the move away from coursework had taken away an important developmental learning process for students – one in which they were able to sink deeply into a subject, to explore it, to refine ideas and to respond to formative feedback from their teachers. The skills developed through coursework were seen as higher than those required for controlled assessment. Skills for controlled assessment were described as more-mechanised responses where students merely memorised something and reproduced it under controlled conditions. In autumn 2011

[51] QCA (nd) *Changes to GCSEs including controlled assessment, information for teachers* (Ref: QCA/09/4174).

[52] Please see DfE (2011) *The evaluation of the impact of changes to A levels and GCSEs: second interim report* reference DFE-RR170 www.education.gov.uk/publications/RSG/NewRsgPublications/Page1/DFE-RR170 for detailed discussion of the impact of the introduction of controlled assessment.

only one head of department commented that controlled assessment kept students very focused all the way through the course, and had thus had a positive effect on learning – although this was a view shared by a small majority of staff interviewed in spring 2011.

Staff in general seemed to feel that there were too many controlled assessments in total across different subjects and that they were taking too much time away from effective teaching and learning time. There was also a concern that students were being over-assessed. The responses where this was raised related specifically to English and history, but there were indications that the concern was more widespread. One head of KS4 felt in autumn 2011 that students were now at saturation level with examinations, tests and controlled assessments, explaining that the centre was trying to establish a culture of assessment *for*, as opposed to *of*, learning.

Students of history and geography were the most positive about controlled assessment, and students of Spanish and French the most negative. History and geography students welcomed being able to use their research notes during the controlled assessment. A majority of the students also stated that they liked the fact that controlled assessment did not take place in the main school hall, which is where they sat external examinations, and that controlled assessment was therefore less stressful.

4.4 Participation and student engagement

4.4.1 Participation

This section reports entry figures for each subject of focus at GCSE level by gender and year, based on JCQ data. Entries fell across all subjects considered (and for all full course GCSEs more generally), although the extent of the fall varied from subject to subject.

4.4.1.1 French

Entry rates in French have continued to fall. Male and female entries are decreasing at similar rates – so there is no reason to conclude that falling numbers are due to a gender influence. The fall in entries between 2010 and 2011 is greater than in recent years – a reduction by 13.4% between 2010 and 2011 compared with 5.9% and 6.1% between 2009 and 2010 and 2008 and 2009 respectively (Figure 23). Given the introduction of the English Baccalaureate and a consequent increased emphasis on modern foreign languages, this is perhaps surprising – it may be that the impact of this change has yet to be seen in qualification entries.

Since 2002 the proportion of female entries has continued to rise. In 2002 females made up 52% of entries; in 2011 females made up 58% of entries.

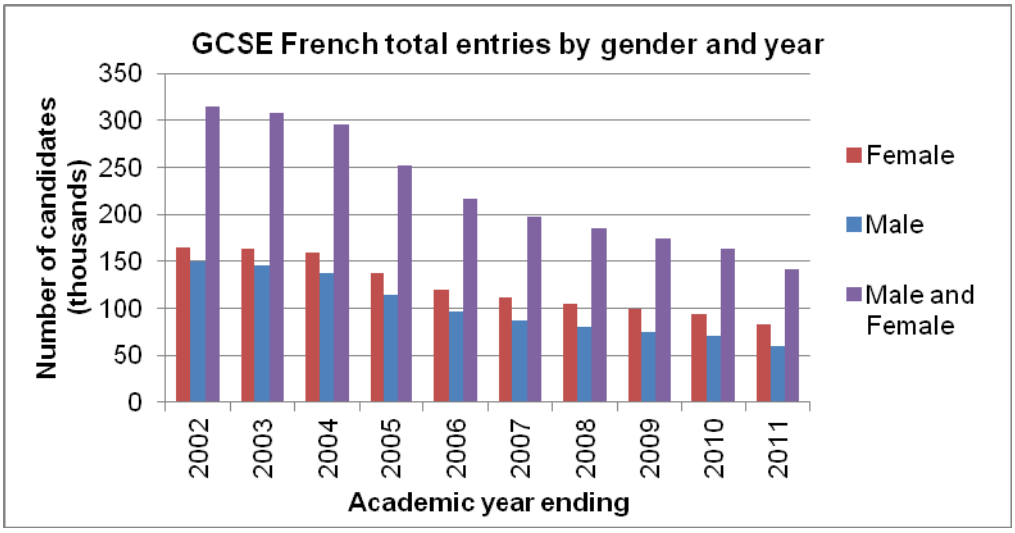


Figure 23: GCSE French entries, 2002–11 (JCQ)

4.4.1.2 Geography

Entries have continued to fall in geography, with male and female entries appearing to decrease at similar rates.

The drop in entries between 2010 and 2011 is greater than in recent years – a reduction by 6.7% between 2010 and 2011 compared with 0.4% between 2009 and 2010 and 3.5% between 2008 and 2009 (Figure 24).

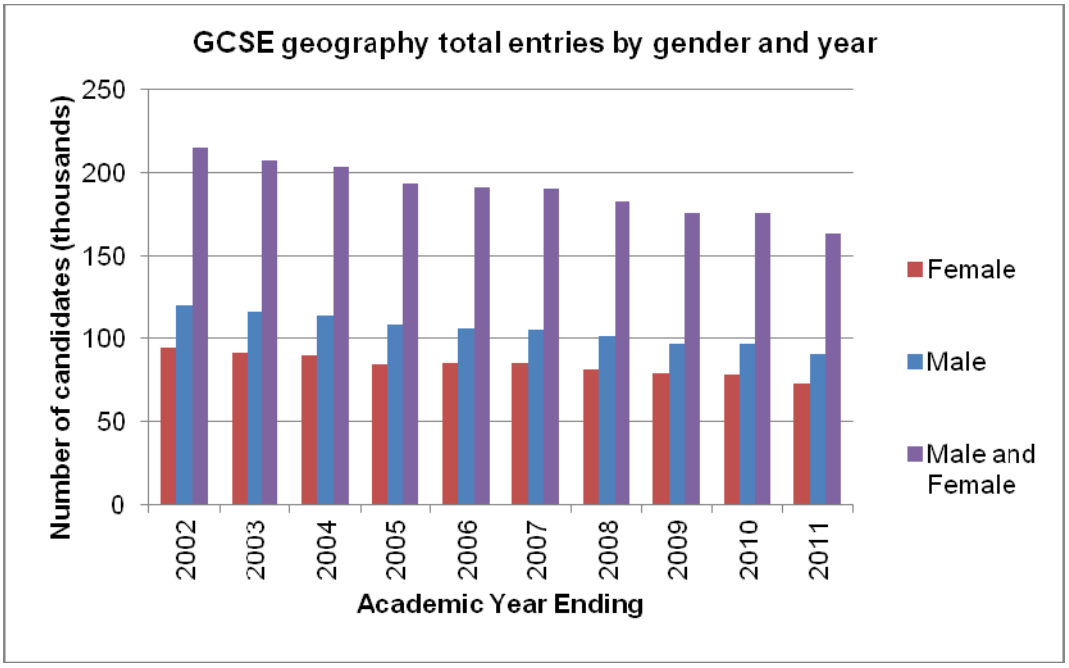


Figure 24: GCSE geography entries, 2002–11 (JCQ)

4.4.1.3 History

In history, entries have fallen slightly since 2010, but remain similar to those seen in recent years.

There do not appear to be any differences in entry trends between males and females: females contribute around 49% of entries every year (Figure 25).

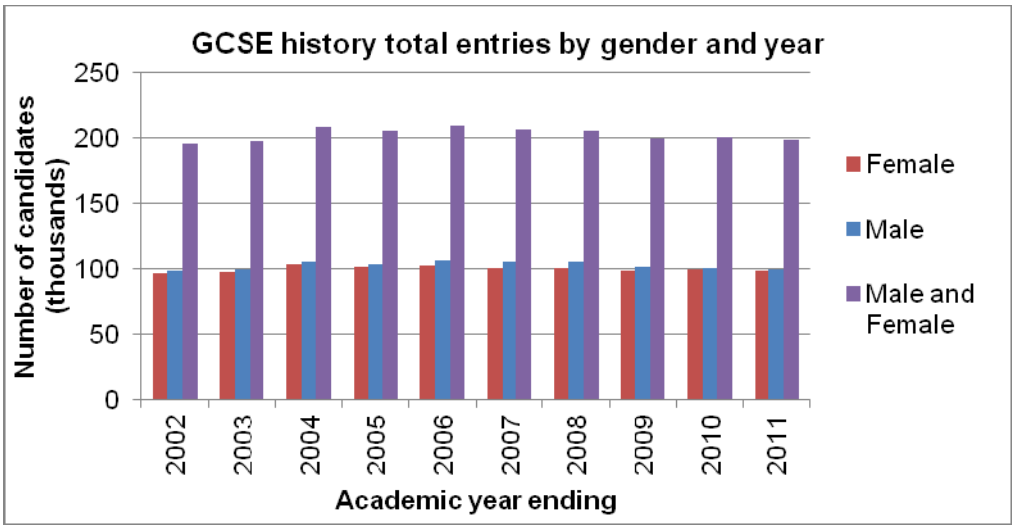


Figure 25: GCSE history entries, 2002–11 (JCQ)

4.4.1.4 Spanish

Entries have fallen in Spanish since 2010. Similar trends can be seen for both male and female entries; females annually make up 58 to 59% of entries (Figure 26).

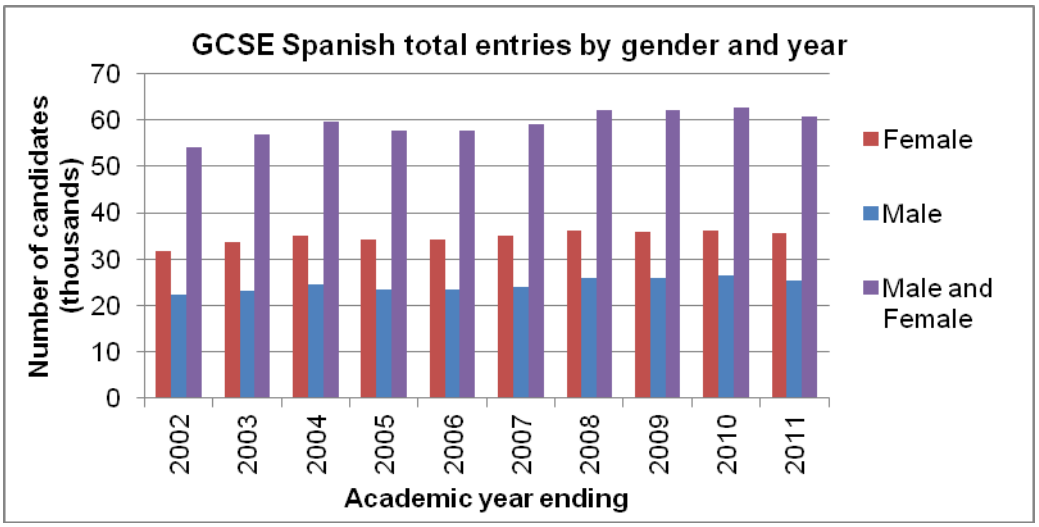


Figure 26: GCSE Spanish entries, 2002–11 (JCQ)

4.4.2 Student engagement

Engagement was usually perceived in terms of focusing on learning to meet the particular demands of examinations rather than a wider commitment to the subject. In the previous

round of fieldwork in spring 2011, the majority of staff interviewed at the case-study centres considered that students were equally or more engaged and motivated than previously, although there were exceptions in the case of modern foreign languages, geography and mathematics. Similarly, most of the respondents to the online survey considered students to be very 'outcome-focused' and aware of what qualifications they needed for progression. Many felt, however, that 'the fun had gone out' of the teaching, because they no longer had time to pursue things that were not directly part of the assessment requirements. Considering content and/or skills that were unlikely to be assessed was seen as an indulgence.

Half the heads of KS4 (interviewed in autumn 2011) had seen evidence of positive change in student engagement that could be related to the new specifications in at least some GCSE subjects. They noted that the introduction of controlled assessment had had a positive impact on students' motivation. On the one hand, students were motivated to do better because they now have small assessments marked and returned by the teacher throughout the year. On the other hand, the lack of specific guidance that teachers can give during controlled assessments makes some young people feel that there is more responsibility on them to get it right. The responses from heads of departments were divided about the significance of controlled assessment as a factor in student engagement, and about whether the impact was the same for students of all abilities. It was given most weight in the cases of French, Spanish and – but only with foundation-tier students – English. These instances apart, the introduction of controlled assessment was not seen as having made an identifiable difference.

Other factors that were identified as having a positive impact on levels of student engagement and motivation were: in geography, the contemporary nature of the curriculum and unitised assessment; in history, the reduction in content, which allowed more time for active learning approaches such as role play, group work and discussions; and, in mathematics, where practical tasks enabled students to link mathematics to something connected to the real world. However, as with other subject responses, there was no uniformity of view among the heads of mathematics. Of those interviewed in autumn 2011, for every one that reported students as being more engaged (particularly the middle- and lower-attaining groups) there was another who felt that students were less engaged because the specification provided little challenge and stimulation for those of higher ability while being too challenging for lower-attaining students.

Students' own views about their level of engagement with, and enjoyment of, the different subjects were also mixed, although predominantly positive. Their comments suggested that they felt that the subject content of French, Spanish, geography and history was up to date

and that the subjects helped them understand the world around them. They reported, however, having issues with the controlled assessment in French and Spanish.

4.5 Progression

Throughout the evaluation, opinion was firmly divided across the six subjects and among teachers, heads of department and heads of KS4 on whether the new specifications provided students with more appropriate skills and knowledge for progression to A level than previously. Interestingly, however, in terms of progression into GCSE in the case of all subjects, one or two centres reported that they had started to introduce new elements (including the development of subject-specific skills) into KS3, in order to prepare students more fully for the GCSE course.

In relation to progression to A level for specific subjects:

- Half the heads of English interviewed in autumn 2011 were of the view that the new specification does not prepare students well for progression to A level study, whereas the rest thought either that there was no difference or that it was too early to make a judgement. Concerns were also expressed, for example, that re-drafting skills were not being developed, and that controlled assessments do not prepare students well for progression to A level because they do not encourage deep learning of subject-specific skills, which are a pre-requisite for KS5. There were, however, some different views expressed in the online survey (conducted in spring 2011), which suggested that several centres believed that GCSE English and English language were likely to prepare students better for A level study.
- Opinion was somewhat less divided on how well the mathematics GCSE prepares students for A level. The majority of heads of mathematics thought that it prepares students less well, or equally as badly, as the previous specifications. In their view, there is not enough high-level work within the GCSE and thus it is too easy to get a very high grade without gaining the skills required for A level.
- Several centres expressed the view that the new GCSE specifications for French and Spanish do not prepare students well for progression to A level, higher education or employment, because there was less emphasis on both subject-specific and generic higher-level skills development. One case-study centre stated specifically that the international GCSE (iGCSE) in French and Spanish prepares students better for A level, because it gives them more opportunities to develop as linguists.

- The introduction of controlled assessment in GCSE history was seen to prepare students well for A level but also to lead to fewer students being interested in taking the subject at GCSE.

4.6 The impact of change on centre behaviour

4.6.1 Patterns of candidate entry

In spring 2011, a large majority of respondents had felt that it was too early to assess whether the changes had had a discernible impact on candidate entry at GCSE (as well as at A level). English and mathematics departments continue to be under pressure to enter some students for GCSE examinations early in order to maximise the number of A*–C grades achieved by the school, but centres vary over which cohort of students are entered early. Of the six mathematics departments interviewed, for example, four entered students early – and of these, two entered the more able, and two the C/D borderline students and those they considered vulnerable.

4.6.2 Management and resource implications for centres

The biggest impact on the centres has been the introduction of controlled assessments and the increased burden this has placed on management and resources. Although controlled assessment is often regarded as a welcome replacement for coursework, there have been difficulties.

QCDA guidance on controlled assessment states:

> *Every school and college must ensure that controlled assessment is managed effectively and operates smoothly to benefit students and comply with the regulations. Schools and colleges should introduce a centre-wide approach to controlled assessment from the start of teaching the revised GCSEs.*[53]

A large majority of centres reported having a centre-wide policy and/or departmental policies in place to manage controlled assessment in spring 2011.

About half of the centres that had in autumn 2010 found the scheduling of controlled assessments a significant added management burden reported in spring 2011 that they were now coping well or satisfactorily with the issue. The remaining half thought that the challenge of fitting controlled assessments into the school calendar was becoming a more serious issue for them, as the number of departments involved was increasing.

In autumn 2011, there continued to be some issues with controlled assessment. Of the three KS4 heads who expressed a view on the management implications of the introduction of

[53] QCDA (2010) *Managing GCSE controlled assessment: a centre-wide approach*.

controlled assessment, two thought that it had had a major negative impact. The key issues they raised were timetabling of initial controlled assessment across different departments, and timetabling for students who had been absent. There were also resource implications: two English and two mathematics departments had had to invest in more copies of books, to accommodate all students undertaking controlled assessment at the same time.

4.6.3 Impact on staff workload

A small majority of centres considered that their workload had increased as a result of the introduction of the new specifications, although few felt that the increase had been substantial. Overall, staff accepted that any changes to examination specifications, such as the introduction of controlled assessment, initially increased teachers' workload. The increase had resulted especially from having to schedule after-school controlled assessments for students who had missed them, and an increase in the number of revision sessions – although the latter was due more to the move from linear to unitised specifications than to the new specifications. A majority of teachers thought that the increase in their workload would be only short term. However, teachers of modern foreign languages reported that they expected their workloads to remain heavier in the longer term, too, because of specific requirements to do with organising and conducting controlled assessment.

5 Summary of findings

The development of a national curriculum, of subject criteria and of examination specifications takes place at a macro-level. But it is how criteria and specifications are interpreted and applied by teachers and individual subject departments in schools that largely determines the pedagogic approach and the learning and interactions experienced by the student. Inevitably, even for the same subject and with similar cohorts of student, perceptions, interpretations and applications differ and change over time. Across the three rounds of evaluation fieldwork undertaken there have been varied and sometimes contradictory views from schools and other stakeholders on the outcomes and impact of the changes to the specifications.

The data from awarding organisations and national statistics provides information about trends in participation and grade outcomes. Data from the case-study centres and stakeholder surveys and interviews capture individual experiences, viewpoints and assessments of the impact of the changes. They reflect differences in ethos, context and what are felt to be the main drivers, and therefore priorities, for a particular education phase.

5.1 Stakeholders' perceptions of the impact of the new A levels on teaching and learning

The content, structure and assessment of A levels have undergone several changes in recent years. Each of the changes has had to address the complicated range of issues related to the role and purpose of A level study and the high-stakes nature of the qualifications for many centres (and for their students, as they compete for university places or other further study) – and this in an educational landscape that is itself complex and shifting.

The views expressed by the various stakeholders are underpinned by individual perspectives on curriculum and qualification development in general, and by what they see as the purpose of A levels and the changes to the specifications in particular. Among the issues commonly highlighted were:

- what constitutes the canon of recognised knowledge for a particular subject, the specialist language and the associated subject-specific skills and methods of enquiry
- how a subject should be taught; how knowledge acquisition is managed in terms, for instance, of the sequence of what needs to be learned and when; subject-specific pedagogy; the management of the learning environment and any perceived hierarchy; and the role of educators
- why a subject should be taught, in the sense of whether it should prepare students for their next steps in life – into further study in an academic or vocational context, or into work – or as part of a democratic right to have the opportunity to be challenged and grow intellectually, and/or to develop the cognitive tools to participate as a stakeholder in society.

The introduction of stretch and challenge in A levels and the reduction from six to four units[54] have been welcomed by the majority of stakeholders. Teachers felt that A2 English literature, geography, psychology, history and French were more challenging, and they considered them better preparation for higher education – in terms of encouraging independent learning, for example. There were some exceptions in terms of specific skills development for a subject – for example, some geography teachers felt that there was not enough emphasis on the research skills required for university study. The views were more mixed for physics, depending in part on where students were progressing to and how the

[54] With the exception of physics, where there continue to be six units in the new specifications.

subject related to their choice of further study. The decrease from six to four units at A level was seen as positive in terms of ensuring greater depth and breadth of study.

Prior to the examination results in summer 2010 for the new-specification A levels, centres had been concerned that there would be a negative impact on students' grades because of what they saw as the greater challenge of some of the new specifications, and because of the uncertainty about how this would be assessed in practice. The cause of this concern, in part, was that centres were not always aware of the role that Ofqual and the awarding organisations had played in ensuring that the first students to take examinations under the new specifications are not disadvantaged. Lack of familiarity with the new specifications and assessments meant that teachers felt less able to predict how the examination questions would assess and reward performance.

A clear distinction was not always made between opportunities for stretch and challenge within teaching and learning – i.e. the higher-order skills developed within the course itself – and how these skills are to be assessed and evidenced in responses to examination questions. Stretch and challenge has also been interpreted in a variety of ways by centres, and there continues to be a degree of uncertainty and ambiguity about what is involved. The stakeholders interviewed had mixed views on the extent to which the new specifications had made a difference – and, if they had made a difference, on the degree to which this had been recognised and the need for change understood. Drawing on findings across the three rounds of fieldwork, however, the majority of centres felt that there was greater stretch and challenge in the new specifications at A2, but that this was not always equally reflected in the assessments, possibly as a result of the different ways in which awarding organisations can introduce stretch and challenge. There was also disagreement between and within subjects as to how best to deliver stretch and challenge and, indeed, the extent to which stretch and challenge was possible for all students.

The structural aspects of the changes to the majority of A levels (i.e. apart from physics, which continues to consist of six units), the organisation of content and the mode of assessment (for example, course work or external assessment) were not necessarily perceived as encouraging greater stretch and challenge. What was important was the effect structural changes had on the teaching and assessment of higher-level skills. Many geography teachers, for example, considered that the removal of the coursework element had reduced the level of stretch and challenge because it had been through coursework that students developed the fieldwork skills required for university study. Centres often discussed stretch and challenge in the context of A levels in terms of developing independent learners. Here the mode of teaching and learning was an important factor, but a high level of teacher

input and initial support was needed if the student was to develop the underpinning skills (such as research skills) required.

The specifications criteria require that A2 assessments must include appropriate demand through the use of a variety of stems in questions, by ensuring connectivity between sections of questions, requiring extended writing in all subjects (except where it had been agreed with the regulatory authorities that this was inappropriate), using a wide range of question types to address different skills, and by including synoptic assessment. Awarding organisations have, accordingly, introduced stretch and challenge in one or more of a number of ways: in the qualification syllabus (for example, by increased content or by placing more emphasis on higher-order skills/concepts), by making changes to the assessment objectives and/or relative weightings, in the paper setting (in the form of amendments to the nature of the questions/tasks), in the marking criteria, or at the grade-awarding process. The different approaches depend in part on differences in the nature of subjects and the way skills and knowledge are organised: for example, in modern foreign languages knowledge, understanding and skills are closely linked, and synoptic assessment promotes stretch and challenge.

Differences of opinion across subjects on whether the ‘application’ of subject skills and knowledge created greater stretch and challenge were largely expressed in terms of the methods of enquiry for a discipline. For example, for physics the greater emphasis on ‘application’ was thought by some teachers to have lessened the more important basic principles of the discipline that come with learning and understanding equations and more-complex mathematical problems. For other centres, ‘application’ had proved too challenging, and they had changed to a different specification that was considered more accessible for their students.

If students were to develop higher-level skills and independent learning, teachers considered that they needed time to adjust their teaching strategies and that students needed time to develop new approaches in their learning. Independent and synoptic learning and understanding are often considered more ‘difficult’ because they may introduce new ways of thinking and ordering knowledge. How much additional time students require to master these ‘higher-level’ ways of thinking about their subject clearly depends on their starting point and what they are used to. Additionally, independent learning approaches need, initially at least, to be scaffolded and supported, which is labour intensive for teaching staff. This was particularly evident in the extended project qualification, which required considerable staff resourcing, but which was also widely reported to be of benefit to student and staff development, with wider stakeholders, including HEIs, recognising the benefits of the qualification. The A* grade at A level is perceived by both teachers and students as

recognising high-attaining students. Some universities now require an A* pass for certain courses.

Although teachers considered coursework a burden, in terms of their workload, they also felt it offered assessment (with a formative and summative element) that supported development. Teachers, managers and students suggested that A level coursework performance was a better indicator of a student's ability than written examinations. Physics apart, the majority of A level students' subject- and non-subject-specific comments suggested that they would wish to retain coursework as part of the assessment regime.

Progression from AS to A2 was considered by many case-study centres to be a larger jump than in the previous specifications. During the 2011 follow-up visits for English literature, history and geography, for example, those who responded thought that the AS was preparing students less well for A2 than before. Analysis of examination data indicated that, for most of the subjects looked at in this study, proportionally fewer students continued to A2 for 2010 completion (new specification) than had continued to A2 for 2009 completion (legacy specification).

The majority of the heads of KS5 reported in autumn 2011 that there was generally greater engagement in the new A levels. Overall, they considered this was less to do with specific content and more to do either with the development of skills or with the structure of the course.

5.2 Grade outcomes at A level

Two important caveats need to be placed on findings from the statistical data:

- summer 2011 data is for entries rather than candidates, which means that the effects of re-sitting may cloud information about overall outcomes
- no data is available yet for candidate characteristics, so any changes in attainment cannot be attributed either to changes in assessment or to the candidature.

In general, entries at A level appear to follow the existing trends and not to have been greatly affected as yet by the introduction of the new-specification awards. For all new-specification A level subjects entries in 2011, entries have decreased very slightly since 2010 (down from 703,000 to 696,000), following four previous years of increasing numbers

since 2007.[55] However, there are variations from subject to subject, with entries increasing for physics, for example.

Considering the new specification A levels as a whole, the plateau effect seen in the A level results in 2010 appears not to have been repeated in 2011. In 2011, the upward trend in average grades that was seen up to 2009 was largely restored. This may have been because the 2011 cohort had higher prior attainment at GCSE than their 2010 counterparts, or because of the introduction of revised guidance for awarding organisations on setting grade boundaries[56] – but this cannot be determined until NPD data is published and analysed during 2012. For the four individual subjects considered, however, there were variations from this overall trend.

The plateauing of grades at A level seen in the 2010 statistical data[57] did not fully reflect the perceived increased level of challenge that had been reported before publication of the results from the summer 2010 examinations. Many teachers had expected to see a drop in grade outcomes after the change of specification, especially at A2. In the event, there were two likely explanations for the plateau in grades: the challenges of the new specification causing the grades to fall, and the impact of measures Ofqual introduced (i.e. placing an effective limit on the extent to which grades should fall) to ensure that the first candidates to take the new-specification A levels examinations were not disadvantaged. This presents awarding organisations with a dilemma: strong guidance from Ofqual to use predictive

[55] Since students of school age (i.e. 16–18) are of most interest in this report, the numbers are based on the SFR data.

[56] The grade boundaries for a particular grade define the upper and lower boundary marks for the grade, for example the boundaries for a grade C might be 47 to 54 marks. Decisions about grade boundaries are made by a committee of examiners and awarding organisation staff which considers script evidence from the current and previous years informed by the available statistical data. The boundaries for the A and E grades are set using this professional judgement approach, with the intervening grades spaced equally in terms of marks between the upper boundary for the E grade and the lower boundary for the A grade. This process is undertaken on a unit by unit basis and then combined, using UMS scores, to provide an overall score and grade for the qualification.

[57] From 1996 to 2009 (the last year in which the previous GCE specifications were awarded), A–level attainment continued broadly on the same improving trend in most subjects – a small steady increase of about 0.06 of a grade per year in grades attained, punctuated by short-term dips downwards from the trend at times of specification change. Also, since around 2003, participation has been increasing in most subjects.

For results in 2010, a year in which another specification change has occurred (the specification change for teaching from 2008 working through to results in 2010), attainment saw a plateauing of grades following a long period of increasing grades – i.e. another shift downward from the trend, reminiscent of results in previous change periods in curricula (although much smaller in size than the dip in 2000/01). Looking at individual grades attained (across all new-specification subjects combined), the same proportion of students obtained the A grade (including those who received the newly introduced A* grade) in 2010 as received an A grade in 2009 (24.6%), whereas, for example, from 2003 to 2009, on average year on year 0.7% more students received an A grade.

matrices[58] for annual GCSE and A level awards based on prior attainment data increases the emphasis at grading on attainment data from the previous year's cohort and so may over time compromise the criterion-related element[59] which currently forms part of grading decisions at A level and GCSE. Additionally, such a move toward grade outcomes comparable with those in previous years would also be likely to prevent GCSE and A level grades increasing in future as a result of improvements in teaching and learning, as such improvements could be interpreted at grading as 'grade drift' (i.e. teachers becoming familiar with the new requirements rather than improvements in teaching and learning *per se*).

Throughout the two years of the new-specification A levels, the difference in levels of performance between females and males has remained largely unchanged. Females consistently achieve higher grades.

5.3 Stakeholders' perceptions of the impact of the new GCSEs on teaching and learning

The evaluation has, for two reasons, had to consider the four GCSEs with specification changes in 2009 (French, geography, history and Spanish) separately from changes to English and mathematics (changed for first teaching in 2010): firstly, because of the different timing of their implementation and the degree to which changes to English and mathematics have 'bedded in' compared with the other GCSEs; secondly, because English and

[58] For example for A level, awarding decisions take account of candidates' GCSE outcomes as a measure of prior attainment to provide predicted likely outcomes of A level results (GCSE outcomes for the current and previous year's A level cohorts are compared). These prediction matrices are used to guide examiners in setting grade boundaries.

[59] Some assessments aim to measure precisely what a candidate can or cannot do, in terms of a set of performance standards. Such assessments are called criterion-related. Other assessments aim to put students on a scale of performance from the strongest to the weakest, without necessarily specifying precisely what each part of the scale corresponds to in terms of performance criteria – these are norm-referenced assessments. The A level and GCSE level awards combine elements of both. Ofqual's work around maintaining standards aims to deal with the choice between prioritising comparable outcomes (essentially a norm-referenced approach) and comparable performance (essentially a criterion-referenced approach) in the first and subsequent years of new specifications:

> The **comparable outcomes** perspective implies that grade boundaries should be fixed so as to take account of any deficits in … examination performance which are unique to the first cohort of candidates. On the other hand, the **comparable performance** perspective entails an acceptance that candidates' results in [the first year of a new specification] should suffer because for this reason they did not produce performances comparable to those which would have been achieved by candidates [in the previous year].

Cresswell, M J (2003) Heaps, prototypes and ethics: the consequences of using judgements of student performance to set examination standards in a time of change. University of London Institute of Education

mathematics are high-stakes qualifications (as a result of school floor-targets), as well as being 'gate-keeper' qualifications for student progression. Assessment of English and mathematics, therefore, is usually the focus of intensive monitoring and involving the deployment of a range of strategies for ensuring students reach the required grade. This is not to maintain, however, that a similarly robust approach is never adopted in relation to other subjects.

Although the updating of course content has increased students' sense of motivation and engagement with GCSE subjects, this positive effect has been somewhat counteracted by the increased focus in teaching and learning on assessment. Generally, with English, there is concern that the amount of content is leading to less depth of study with, in some instances, students no longer being required to read a whole text. Equally, there is a range of opinions on the degree of stretch and challenge for mathematics students of differing abilities.

There were some concerns about unitised assessment for GCSE, but again there were mixed views. Linear assessment was thought by some to promote more in-depth and longer-lasting learning than unitised assessment, particularly in relation to the development of subject-specific skills. Others, however, liked to have the option of unitised assessment for students who learned better by having the opportunity to achieve along the way and to build on previous results.

Some stakeholders considered unitised assessment with opportunities for re-sits and early entry, coupled with pressure on centres and students to get results, to be at odds with the value placed on synoptic assessment and on the development of skills – as opposed to a narrow knowledge-based approach to subjects. The overall consensus from centres and awarding organisations was that there is conflict between the need to gain the 'results' required for school performance targets – using re-sits and early entry to maximise pass rates (e.g. at A*–C at GCSE), which is possible with unitised qualifications – and encouraging learning, development, coherence and a greater understanding of the subject.

A wide range of approaches to controlled assessment[60] is being implemented within and across centres, often depending on the subject being assessed and/or the teacher. Awarding organisations have often interpreted Ofqual's generic guidance differently, meaning that there is sometimes variation in the guidance for the same subject across the different awarding organisations. This has resulted in considerable variation in the amount of

[60] Controlled assessment is a new form of internal assessment of the work of a course, replacing coursework. There is no controlled assessment for mathematics.

support that teachers believe they can give their students in the preparation stage of controlled assessment.

There was concern that some teachers were using strategies, even if often unwillingly, to help students to get the best grade, and that these undermine the validity of the controlled assessments in terms of the knowledge and skills that are meant to be assessed. Examples of strategies include running practice assessments that change little in the 'real' assessments, and modern foreign language teachers setting more controlled assessments than required and choosing the best for each individual student.

About half of the centres that had previously found that the scheduling of controlled assessments involved a significant added management burden reported now that they were coping well or satisfactorily with the issue. The remaining half thought that the challenge of fitting controlled assessments into the school calendar was becoming a more serious issue for them, as the number of departments involved was increasing. The management of controlled assessment continues to be more of an issue for some centres than for others.

It is accepted that certain changes, such as the introduction of controlled assessment, may initially increase teachers' workloads, but that this effect is usually only short term. Teachers of modern foreign languages, however, reported that they expected their workloads to remain heavier in the longer term because of specific requirements to do with organising and conducting controlled assessment.

Most teachers liked the opportunity that controlled assessment offered to choose topics and contextualise the tasks for their students. However, there was no consensus across or within centres and subject areas about the extent to which controlled assessment was an appropriate form of assessment for particular skills and knowledge. There were concerns that, rather than promoting in-depth independent learning, controlled assessment may only test the ability to learn content and regurgitate it. Staff at all levels expressed the need for guidance on best practice in the management of controlled assessment.

Students of history and geography were the most positive about controlled assessment, students of Spanish and French the most negative. History and geography students liked being able to use their research notes during the controlled assessment task. The majority of students also stated that they liked the fact that controlled assessment did not take place in the main school hall, which is where they sat external examinations, and that it was therefore less stressful.

Evidence from many of the case-study centres suggests that the focus on school performance measures is a strong driver for centres to ensure students meet target grades. In many cases this means that teaching and learning is driven almost entirely by the

assessment regime. There is evidence to suggest that, unless specific content and/or skills are expected to be assessed, some teachers and students will consider them not so much a priority as an indulgence. In some centres, however, the focus is less narrow, with centres embracing the change and challenge presented.

Choice of specification was reportedly determined by several factors: the appropriateness for the needs and aspirations of students, consistency with the nature of the subject, and familiarity with the awarding organisation.

5.4 Grade outcomes at GCSE

As for A level, two important caveats are to be placed on the findings for GCSE:

- summer 2011 data is for entries rather than candidates, which means that the effects of re-sitting may cloud information about overall outcomes
- no data is available yet for candidate characteristics.

Nevertheless, on the basis that there is no reason to suspect either major changes in re-sitting behaviour (relating to complete GCSE qualifications, not units) since 2010 or significant differences between the 2011 Year 11 cohort and the 2010 and previous cohorts in terms of trends in prior attainment or other important characteristics, it is possible to draw some tentative findings:

- The overall grade distributions obtained in the new-specification GCSEs in 2011 are a little worse than those obtained in 2010, a slight dip following steadily increasing grades since 2004. While the outcomes for females have plateaued in 2011 compared with 2010, the outcomes for males have fallen since 2010 (by around 0.06 of a grade on average).
- Overall grade distributions have plateaued for all candidates in GCSE French and geography, and decreased a little for GCSE history and Spanish.
- The overall grade distributions of males have got significantly worse for all four new-specification GCSE subjects under consideration: French, geography, history, Spanish. These changes are not seen for English and mathematics, for which specifications had not changed in summer 2011. For the four new-specification subjects, the gap between girls' and boys' overall grades (girls performing consistently better over time in each subject) widened in 2011.

This suggests that:

- The new GCSE specifications are proving a little more challenging for boys than girls, even taking into account attainment differences in previous years of GCSE.

- A similar plateauing effect (slightly more pronounced, in fact – a small dip in GCSE grades in 2011) is observed to that seen for A level in 2010, suggesting that Ofqual's revised guidance to awarding organisations has had the effect of maintaining similar overall grade distributions in a situation where grades might have been expected to get worse.

It should be noted that more-detailed analysis of a wider range of new specification qualifications, coupled with consideration of candidates' prior attainment for the years before and after the specification change, would be needed before this last conclusion can be confirmed.

6 Implications of findings

6.1 Centres' management of change

The timescales for the evaluation have coincided with major changes in education policy following the general election in May 2010 and the formation of the coalition government. The implications of these changes, if not the impact, can already be seen to some extent in centre behaviour and perceptions. Although the impact of the latest round of changes (e.g. the introduction of the English Baccalaureate) is not within the remit of this evaluation, what can be understood is the process of change more generally, and how this is managed in centres. Centres' capacity to manage and embrace a culture of ongoing change is a major factor in assessing the degree and type of impact seen.

Centres need to factor in time to embed change into teaching and learning when they develop and evaluate approaches and build these into centre-wide quality improvement processes. When new specifications and assessments are introduced, many centres reveal anxieties about the amount of content to be taught, the extent to which depth and breadth are required, and how little time they have to make these adjustments. For some centres, the greatest anxiety is about the 'teaching' of content, and this may be the result of either variations in the specifications from different awarding organisations or individual teacher interpretation. A lack of exemplar material or previous examination papers creates concern for many centres, which use such resources to gauge what needs to be taught.

There was evidence across the case-study visits that centres expected workloads to increase at times of specification change but, with a few exceptions for specific subjects, workloads were expected to return to a normal, manageable level following initial implementation and the development of centre-wide processes and procedures.

6.2 The impact of performance measures on centre behaviour and student engagement

A theme throughout the evaluation has been the increasing pressure on centres to adopt, almost exclusively, assessment-driven teaching and learning strategies. Few feel that they can go outside the immediate demands of the examination syllabus, or that they have the time to do so. Re-sit examinations add to the pressures on the teaching timetable, although they may ease the pressure that students feel. The high-stakes nature of A levels and GCSEs – for students as important 'gateway' qualifications and for centres as measures of performance in teaching and learning – results in very outcome-focused teachers and students. Although students appreciate teaching and learning that is 'engaging', student commitment is primarily referred to in terms of 'focus' on gaining the right qualification rather than any wider engagement with a particular subject. Teachers face the daunting task of trying to achieve a proper balance between ensuring that students are well prepared for the examination and putting the work in a broader context so that students enjoy the subject and wish to pursue it further.

6.3 The impact of stretch and challenge

For a number of reasons, the concept of stretch and challenge and how it might best be introduced into public examinations is open to a number of interpretations and approaches. Whether or not teaching, learning and assessment are universally more demanding as a result of the new examination specifications is difficult to judge, for three reasons in particular:

i. The extent to which the new specifications are perceived by a centre to be more or less demanding than previously clearly depends heavily on the centre's starting point. As part of their normal practice, some schools will go beyond what is required by the specifications: they view the examination syllabus as an important part of the GCSE or A level course their students follow – but not the whole of it. Other schools may, for a variety of reasons, stick closely to the examination requirements.

ii. As indicated earlier in this report, greater stretch and challenge may, quite properly, be introduced by awarding organisations in a number of different ways. Some of these approaches will be much more readily apparent to centres than others.

iii. There is some ambiguity, both in the conception and in the regulations, about whether the aim of introducing greater stretch and challenge is to test only the higher-attaining students, or whether it should apply to all students.

6.4 Maintaining standards over time while changing teaching and learning

The findings here raise important questions about whether the revised guidance from Ofqual means that from now on A level grades are to be more heavily influenced than previously by prior attainment. The same question applies if Ofqual continues the grading guidance for GCSE into the second year of new-specification GCSE awards (in 2012).

Guidance from Ofqual would appear to indicate that successive cohorts of A level candidates should have A level grade profiles that are firmly predicated on (or improving at the same rate as) their GCSE grade profile two years previously. From certain points of view, this protects awarding organisations from accusations that the standards applied have artificially changed from one year to the next. It does, however, introduce a form of cohort norm-referencing in another guise. It may negate the opportunity for improvements in teaching and learning to be seen in improvements in grades (which in recent years has been an important feature of more criteria-led A level grading systems), because the awarding processes have a strong assumption that any year-on-year improvements in grades are a reflection of increased familiarity with the specifications and so should be removed. A move towards anchoring GCSE and A level grading patterns on the basis of prior attainment scores would significantly change the meaning of grades in our public examinations. Adopting such an approach entirely would obviate the need for grade-awarding meetings and procedures and would fix the overall pattern of grades for each public examination before any examination papers were sat and marked. Many would see this as a retrograde step, especially in the face of well-articulated arguments that using prior attainment to determine grading decisions is an inappropriate way to ensure the comparability of examination standards over time (Murphy, 2007).[61] As described in the second interim report, however, it should be noted that the Ofqual regulations required awarding organisations to report only if awards fell outside boundaries set by prior attainment – there was no absolute requirement to stay within those boundaries. Their purpose, presumably, was to ensure that consideration of prior attainment grade profiles was given due weight rather than making it the sole factor determining grade-awarding decisions. The extent to

[61] Murphy, R (2007) 'Common test methods' in Newton, P *et al* (2007) *Techniques for Monitoring the Comparability of Examination Standards*. QCA: London.

which it played a part in grade decision making is unknown, but as part of the consultation for the second interim report, Ofqual advised that the awarding organisations reported results outside the threshold range in only a small number of awards.[62]

Further investigation of results for new-specification GCSEs and the second year of new-specification A levels in 2011 will be required, including consideration of prior candidate attainment information, in order to determine whether the apparently similar outcomes at GCSE reflect similar underlying evidence (for example, through considering candidates' prior attainment and demographic information).

The study highlights a number of issues and concerns, some of which have existed for some time:

- The continuing widening of the gap in the standards of performance between male and female attainment at GCSE; whether this is due to females responding more effectively to controlled assessment tasks cannot be determined until a more settled form of controlled assessment is in place. It is also difficult to see how controlled assessment benefits females more than coursework did. What may also be significant here is the form of examinations – the move to a linear examination format may shed some light on the issue. There may also, of course, be factors outside the actual form of the examination itself that need to be taken into account – and this disparity is not a phenomenon that is confined to this country.
- Concerns from centres about the move to linear assessment at GCSE and the impact of this move on lower-attaining students.

One of the themes throughout this evaluation was to consider stakeholders' perceptions of the new GCSEs and A levels and of their impact. In this context it is apparent that it has become increasingly the norm to judge examination effectiveness in terms of how good a preparation it is for what is assumed to be the next step (GCSE, A level courses, higher education). In part, the impact of the new specifications has been to reinforce the view that it is on their predictive validity that examinations should be judged. In the process, what may easily be lost sight of is that national examinations such as GCSE and A level are, or should be, primarily designed to assess how well a body of knowledge and skills has been mastered, and that they may be an end in themselves rather than a stepping stone. As a consequence of the focus on the predictive validity of examinations, there is great pressure

[62] DfE (2011) *The evaluation of the impact of changes to A levels and GCSEs – second interim report*, reference DFE-RR170 www.education.gov.uk/publications/RSG/NewRsgPublications/Page1/DFE-RR170

on, and great attention given to, getting students into the ‘high-stakes’ grade band (say, C and above). Grade distinctions below that tend in practice to disappear.

Ref: DFE-RR203

ISBN: 978-1-78105-079-8

March 2012