Comparative Analysis of A Level Student Work

Final Report

Contract Ref. OF151

Alpha*Plus* Consultancy Ltd

Jenny Smith, Angus Alton, Tom Mitchell

Ofqual/12/5226 March 2012

This report has been commissioned by the Office of Qualifications and Examinations Regulation.

Contents

Executive summary

The Awarding Body Data Archive (ABDA) is a longitudinal project designed to explore the impact of the changes to the structure of GCSE and A level qualifications.[1] One of the areas it was set up to investigate was the impact of the changes to the structure of A levels in terms of the quantity and quality of student work produced in response to the new awards for 2010.

This research project, *The Comparative Analysis of A Level Student Work,* was commissioned by Ofqual as part of the ABDA study. The research was undertaken by Alpha*Plus* Consultancy Ltd over a three-month period between December 2011 and February 2012, using a methodology employed by Ofqual in previous comparative studies. The research focused on evidence of any difference, or similarities, between the 2008 and 2010 specifications and students' performance in the examinations with particular reference to stretch and challenge and the introduction of the A* grade. Research questions developed by Ofqual for the study were:

1. Have the change from six to four units and the introduction of stretch and challenge been effective in improving breadth of knowledge and understanding of the subject?
2. Is there any evidence of stretch and challenge in the responses across the grade levels?
3. How do candidates progress between AS and A2?
4. Is there evidence that candidates can achieve A* without doing the stretch and challenge elements?
5. Has the introduction of A* made it possible to differentiate the most able candidates?
6. What does A* performance look like?

In isolation from other datasets, this initial work can only give an indication of the impact that the changes to A levels have had in practice on the syllabus and student outcomes. It is very important to note that the specifications were selected as likely to highlight an interesting range of subjects and to balance the data collection burden evenly across awarding organisations. It was not the intention of the study to identify individual awarding organisation practice, and for this reason no awarding organisation names are associated with any specification in this report.

Methodology

This project uses a range of methods and includes the analysis of a sample of candidate work and accompanying documents from the 2008 and 2010 examination series collected as part of the ABDA project. The subjects included were English literature, French, geography, media studies, physics and psychology. For all the subjects except English literature a single

[1] The ABDA project is a longitudinal data collection exercise to provide scripts and detailed results data to understand the evolution of specific qualification/subject and 'typical' candidate performance.

awarding organisation supplied the materials for their specification. For English literature, all five awarding organisations supplied materials.[2] This was to allow investigation of the range of ways the changes were effected within a subject as well as across a range of subjects.

The initial phase of the research design focused on developing a set of non-subject-specific stretch and challenge indicators as a research framework to support the identification of stretch and challenge across students' work from different A level subjects.[3]

The approach used for the review of materials was essentially qualitative, building on existing ways of investigating examination materials. It involved a two-stage process. The first stage involved a structured analysis of the assessments, gathering factual information about the exact nature of those assessments and then evaluating them in terms of demand. The second stage involved evaluation of the work produced by candidates in response to those assessments. Subject reviewers were asked to **rate** candidate work in terms of its stretch and challenge qualities, and to put the candidates into an overall **rank** order in terms of their overall level of attainment.

For the **rating** exercise, reviewers were asked to provide a qualitative comment on each candidate's work in terms of the extent to which it displayed the key aspects of stretch and challenge, and then to **rate** each candidate on a 10-point scale, where 0 indicated no evidence of the higher-order skills associated with stretch and challenge.

For the **ranking** exercise, the task for reviewers was to place each candidate into an overall rank order (covering the candidates from both years), again providing a comment to explain which factors were decisive. The reviewers were instructed to base their ranking on a general view of subject ability, using the assessment objectives and published performance descriptions for grades A and E to inform them.

Limitations of the study

Although the exercise has generated a great deal of qualitative data, and it has been possible to subject some of this data to statistical analysis, it is important to note a number of caveats, and as a result to treat the outcomes cautiously.

- **Choice of specifications –** For most of the subjects, only one of the specifications available was reviewed for a given year.
- **Number of reviewers –** Subject reviewers were chosen from Ofqual's panel of subject experts, and there were a maximum of three reviewers per subject.
- **Nature of materials –** Although there was a wide range of scripts reviewed, there were very few at each grade. Any a-typicality in any of these scripts (and a balanced

[2] Where the awarding organisation offered more than one specification in the subject, the materials were supplied for the specification which was closest to the ones supplied by the other awarding organisations.
[3] The indicators developed were: creating connections: identification of related concepts and making comparisons, generalisation, transfer and recontextualisation, rather than seeing topics or skills in isolation; use of reasoning – construction of an argument; use of explanation, application and synthesis of ideas rather than just recall of facts; use of strategies for investigation and problem solving, and understanding, of specialist language and methods of enquiry. An expert group identified by Ofqual, including academics and awarding organisation staff, were invited to respond to an initial questionnaire on the appropriateness of the stretch and challenge indicators.

performance across the units is, of itself, relatively unusual) would inevitably have a large impact on the outcomes in an exercise such as this.

Summary and conclusions

Research question 1

Have the change from six to four units and the introduction of stretch and challenge been effective in improving breadth of knowledge and understanding of the subject?

One of the main findings of this study relates to the introduction of stretch and challenge at the same time as the reduction in assessment (typically from six units to four). The net effect of introducing these two changes concurrently has not been entirely uniform across subjects, but often the reduction in assessment is seen to have worked against the introduction of stretch and challenge. This was not just an issue for specifications that had been reduced to two units at A2. In physics, for example, the removal of the synoptic unit meant that candidates in 2010 had less opportunity to show an overall grasp of the whole course and to show their ability to connect different topics.

Research question 2

Is there any evidence of stretch and challenge in the responses across the grade levels?

The analysis of the scripts suggested that there is a strong positive relationship between A level UMS score and the level of stretch and challenge evidenced within the candidate work, which supports the view that stretch and challenge is being seen at all grades. The qualitative data, however, suggests that for some subjects although stretch and challenge is evidenced in the work of higher-attaining candidates and mark schemes often reward this, some aspects of stretch and challenge are prioritised above others to the detriment, in the view of the subject reviewers, of the study and understanding of the subject as a discipline. There is not sufficient evidence therefore to suggest that the introduction of stretch and challenge has had a widespread positive impact on breadth of knowledge and understanding of a subject, with the exception of media studies and, to some extent, English literature and French.

Research question 3

How do candidates progress between AS and A2?

The pattern of progression between AS and A2 is uneven across the subjects. The gap between AS and A2 is perceived to have **widened** (to varying degrees) in English literature, psychology and French. The gap between AS and A2 is perceived to have got **narrower** (again, to varying degrees) in geography, physics and media studies. These differences are the overall result of structural changes to the qualifications creating more or less demand at AS and/or A2.

Research question 4

Is there evidence that candidates can achieve an A* without doing the stretch and challenge elements?

There are two aspects to this question. First, in subjects such as English literature, media studies and, to an extent, geography, the nature of the assessment consists almost entirely of extended writing. Here, as the banded mark schemes make clear, the stretch and challenge elements are almost impossible to disentangle from those specific to subject knowledge, and it would be almost impossible to score consistently high marks without showing considerable evidence of higher-order skills. In French, similarly, the nature of linguistic development means that it is essentially synoptic; again, it would be impossible to gain very high marks without, for example, the ability to synthesise. In subjects such as physics and, to an extent, psychology, which use a much greater proportion of short-answer questions focused on specific parts of the subject content, the precise theoretical answer to this question depends on exactly how much credit is given to questions meeting stretch and challenge requirements and where the raw grade boundaries are set. This question can be addressed in terms of the reviewers' ratings, which for stretch and challenge correlated very highly with the candidates' UMS scores. Moreover, the ratings for the A* candidates were very high, typically 8, 9 or 10 on the 10-point scale. Thus, the evidence from this exercise is that candidates did not achieve an A* without achieving considerable success with the stretch and challenge elements.

It should, however, be noted that the type of assessment is sometimes at odds with what some reviewers consider the nature of the subject as a discipline. The higher-levels of stretch and challenge available in extended writing assessments in geography, for example, do not *necessarily* reward the best geographers unless they can write essays. In the A level specification reviewed in this study, these higher-order skills appear to be prioritised above fieldwork skills. Any findings therefore need to be considered within the wider context of decision making in relation to the skills, knowledge and understanding prioritised for a subject at this level of study.

Research question 5

Has the introduction of A* made it possible to differentiate the most able candidates?

It has already been noted that the correlations between UMS scores, reviewers' rankings and reviewers' ratings were very high, suggesting that there was a high degree of consensus about the most able candidates, whether viewed in terms of general subject ability or higher-order skills. Further analysis shows that, with a few exceptions, reviewers ranked the A* candidates at the top of the 2010 candidates, irrespective of how they compared with the 2008 candidates. All this evidence suggests that the A* did generally differentiate the most able candidates.

Research question 6

What does A* performance look like?

Findings suggest that the A* generally rewards consistently high performance and evidences both higher-order skills and understanding of the subject at A2. Further data is needed to provide a more robust explanation for some anomalies – for example, where students appear to warrant an A* grade based on the quantitative and qualitative data available but have not been awarded the grade, or where a very few students appear to have been awarded an A* without a high rating for stretch and challenge. Given the small number of reviewers and scripts, however, it is not possible to judge whether anomalies are likely to be systemic or not.

With the necessary caveats, A* candidate performance could be described as displaying many of the higher-order skills effectively within the context of the subject, as characterised by the stretch and challenge indicators:

- creating connections: identifying related concepts and making comparisons, generalisation, transfer and recontextualisation
- constructing an argument
- using explanation, application and synthesis of ideas rather than just recall of facts
- using strategies for investigation and problem solving
- using, and understanding, specialist language and methods of enquiry.

1 Introduction to the final report

The Awarding Body Data Archive (ABDA) is a longitudinal project designed to explore the impact of the changes to the structure of GCSE and A level qualifications. One of the areas it was set up to investigate was the impact of the changes to the structure of A levels in terms of the quantity and quality of student work produced in response to the new awards for 2010. The ABDA project has collated a range of evidence on the legacy and current specifications.

The *Comparative Analysis of A Level Student Work* research project was commissioned by Ofqual as part of the longitudinal ABDA study. The ABDA evidence data is collected in two blocks:

- Activity One: Sample scripts and associated assessment documentation
- Activity Two: Examination data

The research reported here focuses on evidence from Activity One. The aim of this research is to understand the impact of the changes to the structure of A levels under the legacy model and current arrangements focusing on the effects of stretch and challenge, the reduction from six to four units, the introduction of the A* grade, and how these features are embedded in the qualification.

The research was undertaken over a three-month period between December 2011 and February 2012.

1.1 Background to the research

The AS/A2 structure for A levels was introduced in 2000. There have been changes to the specifications since then, with the most recent revisions implemented for teaching in 2008 and first awards of the full new-specification A levels in 2010. The changes were first outlined in the 14–19 Education and Skills White Paper (2005):

- 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 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.

The main changes implemented for A levels (2008–10) as a result of the White Paper were:

- a reduction in the number of units from six to four in the majority of subjects
- changes to coursework weightings in some subjects and the removal of coursework in others
- changes to assessment objectives in some subjects

- the introduction of stretch and challenge in A2
- the introduction of an A* grade.

Mathematics A level is the exception, where the only change at this time was the introduction of an A* grade.

1.2 Research objectives

This project uses a range of methods and includes the analysis of a sample of candidate work[4] and accompanying documents from the 2008 and 2010 examination series collected as part of the ABDA[5] project. The research reported here focuses on evidence of any difference, or similarities, between the 2008 and 2010 specifications and students' performance in the examinations with particular reference to stretch and challenge and the introduction of the A* grade. Research questions developed by Ofqual for the study include:

1. Have the change from six to four units and the introduction of stretch and challenge been effective in improving breadth of knowledge and understanding of the subject?
2. Is there any evidence of stretch and challenge in the responses across the grade levels?
3. How do candidates progress between AS and A2?
4. Is there evidence that candidates can achieve A* without doing the stretch and challenge elements?
5. Has the introduction of A* made it possible to differentiate the most able candidates?
6. What does A* performance look like?

Alpha*Plus* Consultancy Ltd was commissioned to undertake the study, using a methodology employed by Ofqual in previous comparative studies.

1.3 Introducing the concept of stretch and challenge

The original concept of stretch and challenge for A levels outlined in the 14–19 Education and Skills White Paper (2005) was intended to help universities to differentiate between the highest-performing candidates and, through the use of additional extension assessment (AEA) material, offer stretch and challenge opportunities to all students across all types of institution.

In addition, discussions on how to identify the highest-attaining candidates resulted in an A* grade, awarded for the first time in summer 2010 to candidates who achieved a grade A on the A level overall and who also achieved at least 90% or more on the uniform mark scale (UMS) across the A2 units. An analysis of the available literature and policy documents indicates that the two initiatives – stretch and challenge and the introduction of the A* grade – are often conflated. The literature is often ambiguous in terms of how stretch and

[4] The subjects included were English literature, French, geography, media studies, physics and psychology.
[5] The ABDA project is a longitudinal data collection exercise to provide scripts and detailed results data to understand the evolution of specific qualification/subject and 'typical' candidate performance.

challenge is realised in teaching and learning, and also whether assessment should recognise and reward stretch and challenge across all the grade levels or just with the A and A* grade.

In terms of how A levels should be assessed, QCA's advice to awarding organisations in October 2006 was that A level examinations should:

- use significantly fewer structured questions and significantly more open-ended questions, which require the response to be constructed by the candidate
- test understanding and connectivity through synoptic questions
- require significantly more extended writing.[6]

Awarding organisations were specifically 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 (as, for example, in mathematics)
- use a wider range of question type to address different skills – not just short answers and structured questions, but open-ended questions, case studies, and so on
- improve synoptic assessment.

The guidance to awarding organisations stated 'how' A level students should be assessed, which implied rather than defined what was expected in terms of the stretch and challenge evidenced in student performance at A2. With the introduction of a new specification for a high-stakes qualification such as A level, where predictive validity is important for progression pathways, there are inevitable tensions when maintaining qualification standards over time – for example, students experiencing a different challenge in an examination with a new specification should not be awarded a different grade from what they would have received under the old specification.[7]

[6] QCA (2006) Letters to AQA, CCEA, Edexcel, OCR and WJEC from Ken Boston, Chief Executive, QCA.
[7] See Tomlinson, M (2002) *Inquiry into A Level Standards: Final Report,* London: Department for Education and Skills, para 22, December 2002 and Ofqual (2010): 'The prime objectives of maintaining grade standards over time and across different specifications within a qualification type necessarily become more problematic, and engenders more concerns among stakeholders, at times of curricular change' (Ofqual (2010) *GCSE, GCE, Principal Learning and Project Code of Practice*, Coventry: Ofqual, para 6.22). 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, retrieved 19 January 2012.

2 Methodology

2.1 Approach to the research

This study makes a clear distinction between stretch and challenge within teaching and learning and how these skills are elicited within the qualifications and evidenced in student responses. The focus of this study is on the promotion of these skills, but the potential for negative and/or positive impact on teaching and learning will be implicit in the research findings. In order to identify evidence of stretch and challenge in assessment, the underlying assumptions of 'what' was to be assessed needed to be unpicked from 'how' awarding organisations had been asked to develop examination questions.

The initial phase of the research design focused on developing a set of non-subject-specific stretch and challenge indicators (see section 2.2.3.1) as a research framework to support the identification of stretch and challenge across students' work from different A level subjects. The approach used for the review of materials was essentially qualitative, building on existing ways of investigating examination materials. It involved a two-stage process. The first stage involved a structured analysis of the assessments, gathering, first, factual information about the exact nature of those assessments and then evaluating them in terms of demand. The second stage involved evaluation of the work produced by candidates in response to those assessments, considering them in terms of both overall attainment and the extent to which they evidenced the higher-order skills implied by the stretch and challenge initiative and the generic indicators defined for this study.

2.2 Sources and methods of data collection

Data sources for this work included:

- awarding organisation materials for 2008 and 2010
- specification
- question papers and mark schemes
- examiner's report
- any other key documents for evaluating demand (e.g. listening tapes, specification grids)
- a sample of candidate work
- policy documents
- expert group feedback on stretch and challenge paper
- lead reviewer workshop
- expert group focus group.

2.2.1 Materials used in the research

The ABDA project obtained a defined set of examination materials for two years – 2008 and 2010 – across a range of subjects. For the current project, the subjects involved were English literature, French, geography, media studies, physics and psychology. The subjects

were chosen to cover a range of subject types and coursework weightings. Mathematics was not included, because its assessment did not change over the period. For all the subjects except English literature a single awarding organisation supplied the materials for their specification. For English literature, all five awarding organisations supplied materials.[8] This was to allow investigation of the range of ways in which the changes were effected within a subject as well as across a range of subjects.

The materials supplied can best be seen in two parts, syllabus materials and candidate work, corresponding with the two stages of the research. The syllabus materials were the specification, the question papers and accompanying marking schemes and the chief examiner's report. AS as well as A2 materials were supplied. In some cases, not all of these materials were available for both years (see Appendix 1 for a full list of the materials that Ofqual provided to the research team).

In terms of candidate work, awarding organisations were given target A2 UMS scores and asked to supply the complete A2 work of three candidates at each grade (18 candidates in total), including any coursework.[9] A parallel sample of candidates was requested at AS. As with the syllabus materials, not all the samples of work were complete (Appendix 1 provides full details). For the purposes of this project, the scripts were, as far as possible, anonymised and marks removed, so that the experts would be making judgements independently of the candidates' actual outcomes.

2.2.2 Personnel involved in the review

2.2.2.1 Subject reviewers

Ofqual directly recruited three subject experts for each subject. For the purposes of this study, English literature was treated as five separate subjects, with a team of three looking at materials from each awarding organisation. In the event, only two experts were appointed to review psychology and CCEA English literature.[10] One member of each team was appointed lead reviewer, with the task of co-ordinating and summarising the responses of the other members. For English literature, one of the lead reviewers was also asked to take on the role of looking across the outcomes for all five awarding organisations and reporting on key issues raised.

2.2.2.2 Expert group

An expert group with members nominated by Ofqual, including academics and awarding organisation staff, was invited to respond to an initial questionnaire on the stretch and challenge indicators. Responses were received from 6 of the 16 invited to take part (or from an alternative person nominated). In addition, members of the expert group were invited to a focus group with lead subject reviewers to review the initial findings.

[8] Where the awarding organisation offered more than one specification in the subject, the materials were supplied for the specification which was closest to the ones supplied by the other awarding organisations.
[9] For 2008, when there was no A*, candidates were selected on the basis that their A2 UMS score would have qualified for the award.
[10] This was due to lack of availability of subject reviewers

2.2.3 Research approach

Copies of the research instruments used at each stage of the study are included in Appendix 2.

2.2.3.1 Defining stretch and challenge

The range of definitions of stretch and challenge in the literature suggested that a definition that could be recognised across all grades might be problematic. In the first instance, therefore, a set of non-subject-specific indicators was developed (based on the guidance to awarding organisations and the initial literature review) as a research framework to support the identification of stretch and challenge across students' work from different A level subjects and to test the robustness of the indicators within subject contexts.

The indicators developed were:

- creating connections: identification of related concepts and making comparisons, generalisation, transfer and recontextualisation, rather than seeing topics or skills in isolation
- use of reasoning – construction of an argument
- use of explanation, application and synthesis of ideas rather than just recall of facts
- use of strategies for investigation and problem solving
- use, and understanding, of specialist language and methods of enquiry.

A small-scale consultation was undertaken with an expert group identified by Ofqual, including academics and awarding organisation staff, who were invited to respond to an initial questionnaire on the stretch and challenge indicators. Overall, the indicators were accepted as an appropriate 'working definition' for identifying stretch and challenge for this study.

2.2.3.2 Syllabus review

All the subject reviewers attended an **initial briefing meeting** in January 2012 at which the project team explained the background to the project, and the nature of the work required of the reviewers. Here, they also considered the documentation they were going to have to complete and tailored it to the requirements of their own subject. This meeting also provided an opportunity to explore the nature of stretch and challenge in the context of A levels.

Each member of the team then independently reviewed the set of materials provided for each of the two years concerned.[11] For each year they completed a Form A, which consisted largely of factual information about the materials, such as how the content was assessed, the structure of the qualification, the types of question used in the external assessment and so on. Once these forms were completed, the reviewers then completed a Form B. This required the rating of each main factor in the syllabus in terms of demand, and the writing of comments to explain the ratings, especially where they differed across years. As part of this process, the reviewers also rated each external question paper in terms of four factors (**CRAS** – the **C**omplexity of the tasks, the extent to which the **R**esources needed to carry out the tasks were provided to the candidate, the **A**bstractness of the concepts involved and

[11] This part of the work was carried out over a period of approximately three weeks and was home based.

how far the candidates had to generate their own **S**trategy for performing the tasks). These factors have been established as key mechanisms by which the demand of tasks can be raised or lowered, and they relate closely to the ideas of stretch and challenge.

For this part of the work, the reviewers considered the AS materials alongside those from the A2.

Once this stage was completed, the subject reviewers re-convened and the lead reviewer had the opportunity to explore the team's findings with them, and in particular to consider the implications of any major disagreements.

2.2.3.3 Script review, A2

The nature of the research meant that it was important to consider candidate performance in terms of both the extent to which the candidate displayed the elements of stretch and challenge and the overall level of attainment in the subject. To tease these key concepts apart, reviewers were asked to rate the work in terms of its stretch and challenge qualities, and to put the candidates into an overall rank order in terms of their overall level of attainment.

For the **rating** exercise, reviewers were asked to provide a qualitative comment on each candidate's work in terms of the extent to which it displayed the key aspects of stretch and challenge, and then to **rate** each candidate on a 10-point scale, where 0 indicates that there was no evidence of the higher-order skills. Here, the ratings given to the benchmark work would be a crucial guide in the process. For the **ranking** exercise, the task for reviewers was to place each candidate into an overall rank order (covering the candidates from both years), again providing a comment to explain which factors were decisive. The reviewers were instructed to base their ranking on a generalised view of subject ability, using the assessment objectives and published performance descriptions for grades A and E to inform them, For this part of the work, reviewers were not permitted to give any pair of candidates equal ranking.

At the start of this phase of the work, the subject reviewers attended a further meeting (the **cross-moderation meeting**), for briefing and cross-moderation purposes. They were provided with hard copies of six candidates' work and asked to work together to **rank** and **rate** these scripts. The scripts could then be used as benchmarks for the rest of the process. Reviewers were then provided with all the remaining candidate work –again in hard copy – and worked independently at home to complete the rating and ranking process.[12] The experts recorded all their comments and their rankings and ratings on a spreadsheet provided for the purpose.

2.2.3.4 Meeting to review findings at A2

Once the process was completed, the teams of subject experts reconvened (at the **lead reviewers' meeting**) to come to a view on the effects of the various changes between 2008 and 2010, taking into account evidence from both the assessment document review and the

[12] Where all the materials were available, this involved a further 30 candidates, to produce an overall rank order for all 36. This means there is a very substantial body of evidence to bear in mind to complete such a process.

review of candidate work. The lead reviewers then reported on the key issues raised, to explore the extent to which there was any overlap across the different subjects.[13]

2.2.3.5 Script review: AS

A key drawback of the AS work supplied (though it was drawn on essentially the same basis as that of the A2 material), in terms of the research question about progression between AS and A2 was that it was not the AS work of the candidates who were considered at A2.[14] A somewhat different approach was therefore taken to the ranking of the AS work, with the experts being provided with all the AS work, on-line, and in rank order of their UMS mark; they were asked to compare this work with the A2 work in their own rank order, and to provide a general, qualitative report on performance at AS in the two years. They were asked to focus, in particular, on any differences between the two years, in terms of performance at AS and at A2, and on the extent to which the AS work displayed any of the higher-order skills associated with stretch and challenge.

2.2.3.6 Focus group

Once the subject experts had completed their task, the lead reviewers attended a **focus group meeting** with some of the assessment experts who had responded to the original paper on stretch and challenge. In this meeting, the lead reviewers outlined the key issues raised in their subjects and the assessment experts were invited to comment on these and other relevant issues, using the key research questions to structure the discussion.

2.3 Limitations of the study

Although the exercise has generated a great deal of qualitative data, and it has been possible to subject some of the data to statistical analysis, it is important to note a number of caveats, and as a result to treat the outcomes cautiously.

Choice of specifications – For most of the subjects, only one of the specifications available was reviewed for a given year. The way the ABDA project was set up to some extent mitigates this, in that it was set up to identify the largest specification in the market place and then to identify comparable specifications from other awarding organisations over time. However, it would be dangerous to assume that the particular specification under review was typical of the way the subject was considered by other awarding organisations. Even with English literature, several of the awarding organisations offer more than one specification, with the second one by definition distinctive from the first.[15] Restricting the choice of specifications was of course necessary from a practical viewpoint, but it is important to bear in mind that it can be dangerous to assume that findings for one specification would apply to another.

Number of reviewers – Subject reviewers were chosen from Ofqual's panel of subject experts. Most have long experience of involvement in evaluative exercises of this type, and a further factor in their selection was that they had no personal connection with the awarding organisation whose specification they were evaluating. Given the practicalities, there were a

[13] Here, only the co-ordinating lead reviewer reported on English literature, to avoid unbalancing the process.
[14] In practice, obtaining representative candidates who score consistently across AS and A2 is problematic.
[15] It is a regulatory requirement where awarding organisations offer more than one specification in a subject that the two should be clearly distinctive from each other.

maximum of three reviewers per subject. This is a small sample and therefore potentially unreliable. Table 1 shows the correlations between each reviewer's rankings for the student work. As the ranks are ordinal in nature (i.e. they are necessarily ordered, but the distance between their values is not meaningful), Spearman's rank order correlation is used. As Table 1 shows, generally there was agreement between the reviewers, with the exception of English literature (4) where reviewer 2 was not closely in line with the decisions of the other two reviewers. There were higher correlations in 2010 in French, and slightly higher correlations in English literature (1) and physics. There were lower correlations in 2010 in geography and English literature (2). The differences in opinion are reflected in the discussion in the findings section (Section 2) and are pertinent in relation to the amount of change seen in students' work.

Table 1: Correlation between subject reviewer ranking by year and subject

			Spearman rank correlation[16]		
	Reviewer A	**Reviewer B**	**2008**	**2010**	**Overall**
French	1	2	0.68	0.92	0.85
	1	3	0.86	0.91	0.88
	2	3	0.85	0.91	0.88
Geography	1	2	0.94	0.77	0.86
	1	3	0.95	0.55	0.76
	2	3	0.94	0.48	0.72
Media studies	1	2			
	1	3			
	2	3		0.75	
Physics	1	2	0.83	0.93	0.87
	1	3	0.99	0.99	0.99
	2	3	0.85	0.94	0.88
Psychology	1	2	0.96	0.97	0.96
	1	3			
	2	3			
English literature (1)	1	2	0.77	0.92	0.84
	1	3	0.90	0.98	0.95
	2	3	0.80	0.88	0.83
English literature (2)	1	2	0.91	0.75	0.81
	1	3	0.95	0.82	0.88
	2	3	0.96	0.70	0.87
English literature (3)	1	2	0.93	0.92	0.91
	1	3	0.84	0.81	0.81
	2	3	0.92	0.87	0.89
English literature (4)	1	2	0.16	0.61	0.39
	1	3	0.80	0.88	0.84
	2	3	0.31	0.48	0.36
English literature (5)	1	2		1.00	
	1	3			
	2	3			

[16] For example, the correlation between Reviewer 1 and Reviewer 2 in French is 0.68 for the2008 candidate work and 0.92 for the 2010 candidate work. Columns titled Reviewer A and Reviewer B indicate which reviewers are being compared.

Please note that in the table a figure of 1.0 would indicate a perfect linear agreement, a figure of –1 would indicate a negative correlation (i.e. the higher the UMS score, the lower the reviewer ranking seen), and a figure of 0 would indicate no identifiable relationship between UMS score and reviewer ranking.

Nature of materials – Although there was a wide range of scripts reviewed, there were very few at each grade. Any a- typicality in any of these scripts (and a balanced performance across the units is, of itself, relatively unusual) would inevitably have a large impact on the outcomes in an exercise such as this. There were also some gaps in the evidence base, across the full range of subjects, and on occasion some samples of candidates' work were incomplete. For media studies and English literature (5), there was no work from 2008. The exercise looking at French was particularly constrained: the limited oral work available made it difficult to match what was provided with the other units, so a separate, and very limited, ranking and rating exercise had to be carried out on the orals.[17]

Nature of the numerical data provided – It was not possible to access comparable background data on the candidates across all specifications. In many cases, both the A2 and total A level UMS scores were provided; in several, only the total A level score was provided; in one, only the A2 UMS score was provided. Moreover, because of the nature of the anonymisation, it was impossible to derive both sets of scores where one was missing. With the one exception, correlations are based on the total A level UMS score. However, in addition to this one exception, it is worth noting that in some ways it would be preferable to use the A2 score, since differentiation of A* from the A grade is also based on that score, and the stretch and challenge elements are likely to be more concentrated in the A2 units.

2.4 The scope and limitations for this report

This study is part of the ongoing work of the ABDA programme and is restricted to the size and sample of candidates' work that has been archived from 2008 and 2010 (the scope of the candidate work available is included in Appendix 1). It should also be recognised that the findings here are discussed largely in isolation from any analysis of the examination statistical data[18] and any background data on candidates and centres, or the teaching and learning. It should also be noted that candidate work from 2010 is from the first year of awarding for the new-specification A level. Candidates and teachers are therefore likely to be less familiar with the specifications and assessments, and this could affect candidate performance in the examinations.

This initial work gives an indication of how the changes to A levels have had an impact in practice on the syllabus and student outcomes. It is very important to note that the specifications were selected as likely to highlight an interesting range of subjects and to balance the burden evenly across awarding organisations. The study is not intended to identify individual awarding organisation practice, and for this reason no awarding organisation names are associated with any specification in this report. The assumption is that roughly the same conclusions could be drawn from any subject/specification considered, and the review of all English literature specifications was intended to test this assumption.

[17] There were two oral units available for AS at 2010 and four for A2 at 2010.
[18] Triangulation of findings from this study with the statistical data will commence in April 2012.

3 Research findings

Stretch and challenge is reported in this section in terms of the stretch and challenge indicators developed (see Methodology, section 2.2.3.1):

- creating connections: identification of related concepts and making comparisons, generalisation, transfer and recontextualisation, rather than seeing topics or skills in isolation
- use of reasoning – construction of an argument
- use of explanation, application and synthesis of ideas rather than just recall of facts
- use of strategies for investigation and problem solving
- use, and understanding, of specialist language and methods of enquiry.

The research findings are discussed in the following sections:

- main changes in the subject specifications in 2010
- stretch and challenge evidenced across the grade profiles
- stretch and challenge in relation to breadth and depth of subject knowledge, skills and understanding
- the relationships between specification and assessment
- the A* grade in terms of stretch and challenge and mastery of subject/discipline
- progression from AS to A2
- the extent and impact of variation observed within subject specifications across awarding organisations for English literature.

3.1 Main changes to specifications in 2010

The main changes seen between 2008 and 2010 in the specifications for each subject are outlined in Table 2. A summary of the findings for each subject can be found as an appendix to this report (see Annexes 1–6).

Notes on Table 2:

- Although the number of units in physics nominally stayed the same, the 2008 syllabus had two split units, one at AS and one at A2. Each comprised an external written paper and an assessment of investigative and practical skills, so that there were eight discrete pieces of assessment. The assessment objectives in physics saw the loss of a specific objective for synoptic assessment and significantly greater explicit reference to how science works.
- In the 2008 physics A level, one AS sub-unit (weighted 7.5%) and one A2 sub-unit (weighted 5% of the A level) offered the option of coursework or a practical examination. In 2010, one full unit at AS (weighted 10% of the A level) and one at A2 (weighted 10%) assessed practical and investigative skills through a combination of an externally set, internally assessed assignment and an internal assessment of practical skills.

- The 2008 assessment objectives in English literature were further complicated by being slightly different at AS and A2, as well as differently weighted. The greater number of texts to be covered in English literature involved greater expectations of wider reading for context and comparisons, and there was a subsequent reduction in emphasis on close reading.
- The 2008 specification for media studies was not available for a clear statement about the assessment objectives and their weightings.
- In 2010, psychology was re-classified as a science, so had to conform to the science criteria. The assessment objectives were thus considerably changed, although not in number.

Table 2: Main changes seen in 2010 for specifications reviewed

	Number of units	Number of assessment objectives	Coursework weighting	Content	Other
English literature	6→4	5→4	30%→40%	Increased in terms of number of texts to be studied	More open-book examinations Loss of a unit focused on synoptic assessment
French	6→4	4→3	15% (optional) → none	Change in the way topics are used and in the number to be studied	More-complex assessment to ensure coverage of assessment objectives
Geography	6→4	4→3	22.5% (some optional)→ none	Reduction in number of topics	Loss of coursework meant fieldwork could be assessed only indirectly Loss of a unit focused on synoptic assessment
Media studies	6→4	?[19]→4	40%→50%		Greater emphasis on technology
Physics	6→6	4→3	Completely revised: see notes	Broadly the same	Loss of a unit focused on synoptic assessment
Psychology	6→4	3→3	15%→none	Reduction in number of topics covered	Absence of coursework meant assessment of methods of enquiry theoretical rather than through a personal study. Loss of a unit focused on synoptic assessment

[19] Subject criteria and specification for media studies were not available for 2008.

3.2 Stretch and challenge evidenced across the grade profiles

Half of the members of the expert group who responded to the stretch and challenge questionnaire considered that stretch and challenge should be evidenced at all grades; this response was, however, always qualified in some way.[20] The analysis of the scripts suggested that there is a strong positive relationship between A level UMS score and the level of stretch and challenge evidenced within the candidate work, which supports the view that stretch and challenge is being seen at all grades. Table 3 shows the level of correlation between A level UMS score[21] and the mean subject reviewer rating for stretch and challenge.

Table 3: Pearson correlation between A level UMS scores and the mean subject reviewer rating for stretch and challenge

	2008	2010	Overall*
French	.862	.974	.885
Geography	.858	.863	.822
Media studies		.740	
Physics	.879	.928	.895
Psychology	.908	.923	.890
English lit (1)	.803	.941	.870
English lit (2)	.940	.895	.902
English lit (3)[22]	.942	.922	.899
English lit (4)	.898	.779	.834
English lit (5)		.899	

*This is the correlation between the UMS score, as a proportion of 1, and average stretch and challenge rating. The two year columns consider the correlations for candidates in that year, whereas the overall column considers all candidates regardless of year.

There is also a strong positive correlation between the A level UMS score and the mean subject reviewer ranking for all scripts.[23] This is in line with expectation – reviewers were asked to put the candidates into an overall rank order in terms of their overall level of

[20] One respondent thought that the higher-attaining students were more likely to provide evidence of stretch and challenge. Another felt that each indicator signified a 'spectrum of a class of attainment' rather than a dichotomy; context, audience and level of complexity were considered relevant variables. The inverse of the descriptors (e.g. 'unable to explain any ideas' or 'can't use any methods of enquiry') was thought unlikely to justify any grade at all. A third respondent felt that the ability to question facts, to discuss answers rather than just get the right answer was an important part of teaching and learning for all students as well as of what is examined.
[21] A2 UMS score is not available for all specifications.
[22] This data is based on A2 UMS as A level UMS is not available
[23] Note that reviewers did not know the UMS score or grade when ranking or when rating for stretch and challenge.

attainment, and so the rank order would be expected to correlate strongly with UMS. When the data for A level UMS score and stretch and challenge mean rating[24] was analysed separately for 2008 and 2010,[25] the UMS score was found to be highly significant in predicting stretch and challenge rating for all subjects.

In terms of differences between stretch and challenge between 2008 and 2010, year was found to be a significant predictor of stretch and challenge rating only for psychology, French and geography (see Figures 1–3 below). In the case of psychology this was because, for a given UMS, the stretch and challenge rating tended to be slightly higher in 2008 than in 2010, especially in the lower half of the mark range, whereas in geography the opposite was true, again particularly in the lower half of the mark range. In French, the outcome arose because, for the 2008 candidates in the lower half of the mark range, there was very little correlation between UMS and stretch and challenge rating, while, for the 2010 candidates, the correlation was strong throughout the range.

What is particularly noticeable about all three results is that it was lower-attaining candidates who were affected. The UMS marks and stretch and challenge ratings for higher-attaining candidates correlated well in both years for all subjects.

[24] Mean average of reviewers' rates for a particular subject.
[25] For each subject, a linear model has been fitted with stretch and challenge rating as the dependent variable (i.e. the outcome) and UMS score and year as the independent (i.e. the predictors). No models were fitted for CCEA English literature and OCR Media studies as no 2008 data was available for these, and no model was fitted for WJEC English literature as there was no UMS data available.

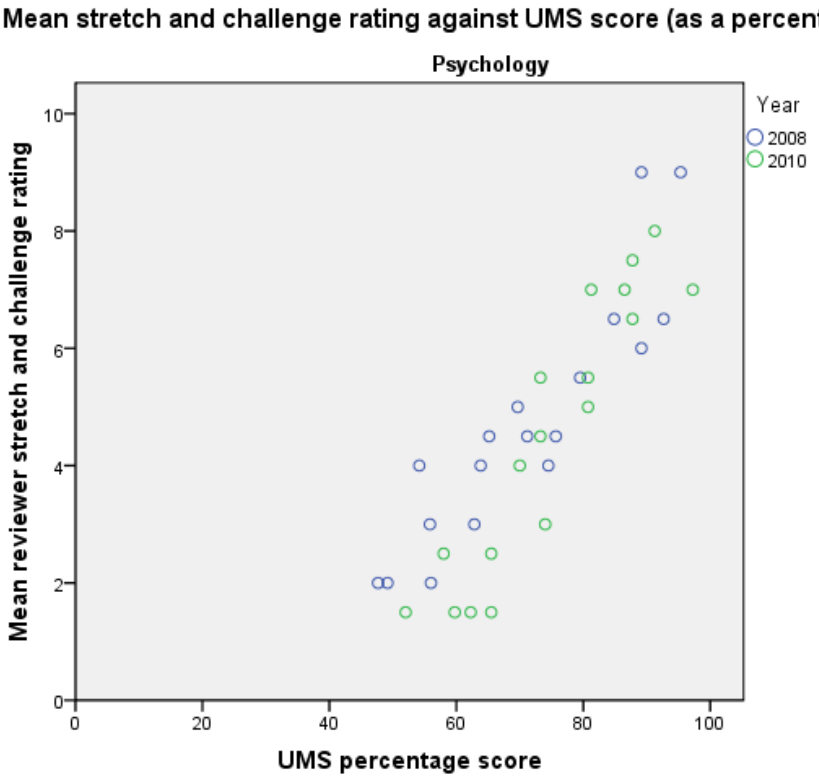

Figure 1: Psychology – mean stretch and challenge rating against UMS score (percentage) by year

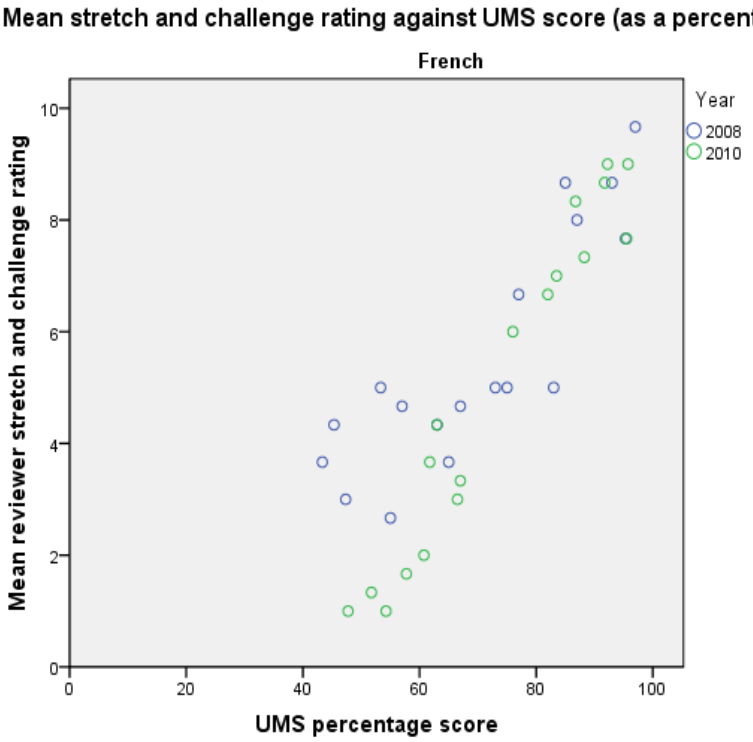

Figure 2: French – mean stretch and challenge rating against UMS score (percentage) by year

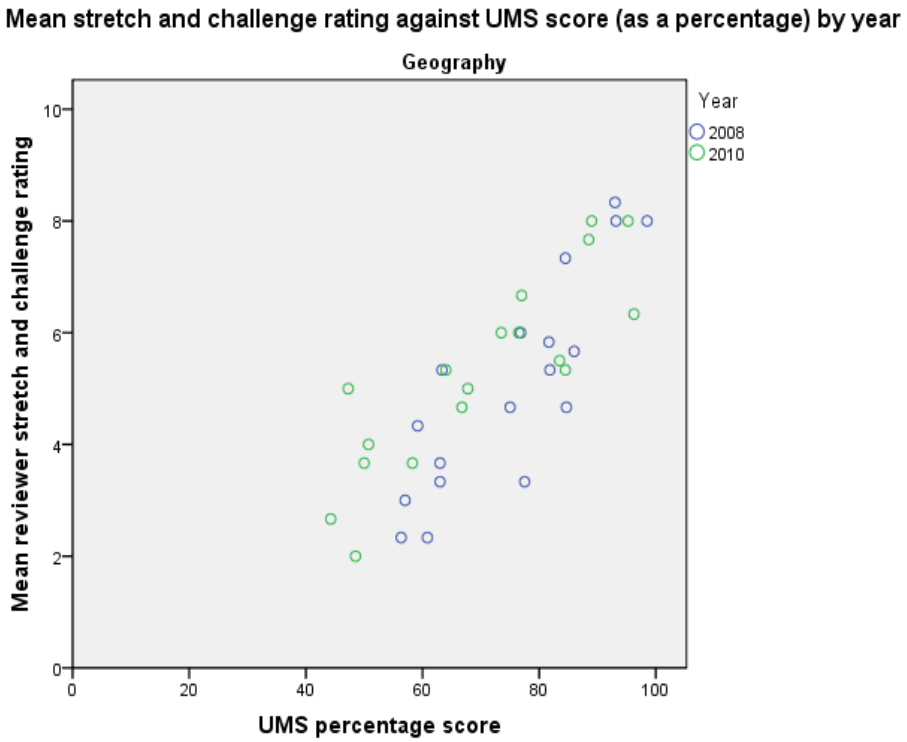

Figure 3: Geography – mean stretch and challenge rating against UMS score (percentage) by year

There was also a very strong correlation between the subject reviewers' own rankings and their stretch and challenge ratings, typically well over 0.9. Interestingly, even for the lowest-ranked candidates, it was relatively unusual for their stretch and challenge rating to be below 3 on the 10-point scale used. This was true for both 2008 and 2010 candidates. It clearly suggests that the relevant skills are present in the work of almost all candidates, while the extent to which they are displayed depends strongly on the overall level of subject knowledge and understanding.

Scatter diagrams showing grade, year and stretch and challenge for all subjects can be found in Appendix 4.

3.3 Link between stretch and challenge and breadth and depth of subject knowledge

Often the definitions of stretch and challenge in the literature seem to link it inextricably to breadth and depth of subject knowledge and skills. Feedback from the majority of the expert group suggested that the generic indicators were likely to be evident in all subjects, but there was a sense that this might mean different things in different subjects and might 'encompass subject knowledge as well as applying higher-level skills'. Evidence from the review of the awarding organisations' documentation and the candidate work suggests that between 2008 and 2010 the changes to assessment objectives and weightings have not necessarily meant that there is more or less stretch and challenge – just that it may be different. Any change in depth or breadth of subject knowledge was largely in terms of a change of emphasis, between content and skills, for example.

In French, for example, there is a tension when defining breadth and depth of subject knowledge, and the removal for 2010 of an assessment objective focusing on study of the culture was thought to emphasise skills over content. While it could be argued that depth and breadth is primarily knowledge and application of the language, it could equally be considered that breadth and depth should include knowledge of the culture and study of its literature (for example) as part of the subject as a discipline. Although some opportunities for stretch and challenge could be seen to have been lost (creating connections, for example), however, stretch and challenge was felt to be embedded within the synoptic element of language anyway e.g. the spoken and written word. In 2010 candidates had to show their linguistic skills at a greater level.

There were some tensions evident in the 2010 examination papers between the assessment of stretch and challenge and the breadth and depth of subject knowledge. In geography, for example, how some aspects of stretch and challenge were assessed – e.g. longer essay-style questions – was considered to be at the expense of identifying depth and breadth of subject skills, knowledge and understanding. The removal of coursework likewise meant that students were not assessed on their actual fieldwork skills but on their ability to evaluate aspects of it in the examination (e.g. the risk assessment). Overall, the subject reviewers commented that they did not feel that they 'knew the 2010 candidates as well as geographers' compared with the 2008 candidates. In terms of stretch and challenge, although the essay-style questioning attempted to promote synoptic assessment in 2010 (creating connections), it could be argued that this was at the expense of the stretch and challenge that relates directly to the subject as a discipline (use of specialist language and methods of enquiry). Geography was the most problematic of the specifications in terms of allowing for differentiation at the higher grades, a point that is reflected in the findings from the statistical analysis of the ranking, rating and UMS scores, with less agreement seen in decisions between the reviewers.

In 2008 physics candidates earned marks roughly in proportion to the amount of explanation and calculation in the questions themselves, although there was some variation, particularly where the total marks were low. In 2010, candidates usually scored a greater proportion of their marks in calculations when compared with the proportion in the questions, as suggested by comments in the examiners' reports such as 'Answers to numerical sections were usually approached much more confidently, and completed much more competently, than those to parts requiring description, explanation, or even straightforward recall'. This loss of a whole assessment objective (AO4: synthesis of knowledge, understanding and skills) was felt to make the overall assessment easier and was likely to lead to less differentiation at the top end of the ability range. Marks for AO4 could provide a cushion where candidates demonstrate knowledge without the higher-level skills of evaluation and analysis, which particularly effected middle- and lower-attaining candidates. The questions for the research-based essay in 2010 are more general and provide less structure for extended responses which the subject reviewers considered more challenging for the whole range of candidates.

3.4 Compression in assessment of subject knowledge

One of the main findings of this study relates to the introduction of stretch and challenge at the same time as the reduction of units (typically from six units to four). Introducing these two

changes concurrently has had an effect on all subjects. Although these effects have not been the same in each subject or specification, often the elements of stretch and challenge were judged to have been compromised by the reduction in assessment.

In French in 2010, the reduction in the number of units has in some ways increased the stretch and challenge in that many of the same skills are assessed in a reduced number of units. In some senses, therefore, the assessment has become more complex, with candidates needing to demonstrate multiple skills in particular questions. The external assessment of this specification means that the emphasis is on productive skills, which affects middle- and lower-attaining candidates as there are fewer opportunities to achieve and candidates have to think much more on their feet throughout. For higher-attaining candidates there are fewer opportunities to demonstrate their knowledge and receptive skills. There is also a greater pressure on time as all the skills and knowledge except for speaking have to be assessed in a single unit; as a result, assessment activities that are actually very similar across the two years, such as the discursive essay, assume a larger proportion of the whole assessment. The reduction to two units also means that there is a reduction in the breadth of knowledge, as one aspect of study for the culture and society of the country is required as opposed to two in 2008 for either option.

The tension between stretch and challenge and the opportunities to demonstrate knowledge and skills was not just an issue for specifications that had been reduced to two units at A2. The removal of the synoptic unit meant that physics candidates in 2010 had less opportunity to show an overall grasp of the course and to show their ability to connect different topics. In geography the range of unit options was thought to lead to fewer opportunities to make links across the different elements of the course. This made synoptic assessment across the course more problematic.

Across both years in English literature there are opportunities for synoptic assessment implicit within the skills of the discipline (e.g. comparing and contrasting texts and thematic approaches to textual analysis). There is less evidence in the specifications for 2010 of a requirement for close textual analysis in terms of form, structure and language. Typically, this has been replaced by an emphasis on thematic study, breadth of reading, reader response and tasks which are rooted in comparison and contrast between texts. This change may appear to diminish the notion of a demanding set text examination as a precursor to advanced literary study, but it may also allow candidates to achieve a greater understanding of the subject as a whole. For some, but not all, of the English literature specifications reviewed, the changes to the assessment objectives, the reduction in the number of units and a focus on wider reading and thematic study have arguably led to an improved performance in the examination papers and more evidence of candidates employing the kind of skills associated with stretch and challenge.

The subject reviewers reported that, in aspirational terms, maintaining a wide breadth of study and choice is a sensible endeavour but one that could be narrowed down by deploying some of the reductionist approaches permitted by the specifications.

In English literature, geography, physics and psychology there was the loss of a unit focused on synoptic assessment.[26] Most reviewers saw this as a retrograde step because, although in theory for 2010 synopticity was intended to be embedded across A2, there were issues identified in practice (geography and physics are examples where looking across the subject and making connections became much more limited in form and degree). For the majority of subjects, however, there was little option but to remove the synoptic unit, given that there are only two units at A2 and the assessment of other skills would have to be squeezed into the other unit.

3.4.1 Reframing or classification of the subject

For the majority of the subjects, with the exception of English literature, there had been some reframing/reclassification of the subject as a discipline at A level in the 2010 subject criteria. In physics there has been a move (already discussed above) to an emphasis on explanation and away from mathematical skills. The increased emphasis in French on linguistic skills rather than close study of the culture has also already been discussed. In geography the emphasis is much less on a more science-based process and much more on an approach to 'softer' values and attitudes.

A key change for media studies is for candidates to be taught how to use 'appropriate media facilities and technologies' 'before embarking upon assessed work' and the need for 'adequate software, equipment and staff training'. The need for the last two points has major implications for the delivery of the syllabus. Conceptually, and in terms of practical work, the demand of the A level has increased considerably from 2008 to 2010.

One of the main changes in the 2010 specification is the introduction of the notion of psychology as a science. The 2010 psychology specification states that it has been designed to provide a broad introduction to the scope and nature of psychology as a science, bringing the content up to date while at the same time retaining the existing features of the previous psychology specification. This in its turn brought the expectation that psychology should conform to a number of set criteria, identified as 'How Science Works'. This change in identity has not led to significant changes in demand in terms of content requirements. In fact, the majority of criteria identified under How Science Works were met in the 2008 specification through the coursework and research methods elements.

3.5 The relationship between specifications and assessment

For the majority of specifications there is less assessment time in 2010 and often different types of question that may or may not offer what is considered to be sufficient opportunity to evidence stretch and challenge and/or subject knowledge and skills within the mark allocation and overall time frame for an examination paper. Concerns were raised by a number of the subject reviewers that this had led to a reduction in the content and skills being assessed.

In some subjects, such as French, less assessment means that there is less predictability, which makes the assessments more demanding. For other subjects, such as geography, the lack of variation in question types, in particular the over-reliance on extended writing

[26] See Table 2 in section 3.1 for a summary of the main changes seen between 2008 and 2010 in the specifications for each subject.

questions, increases predictability because there are fewer topic areas and skills that are suited to the question style.

One of the key findings of this study, therefore, is that the assumption that an increase in questions requiring extended writing will identify greater stretch and challenge and breadth of subject knowledge and skills has not always been realised in candidate performance. Equally, however, short-answer questions, while effective in assessing certain types of subject knowledge, do not necessarily offer opportunities for candidates to evidence the higher-order skills associated with stretch and challenge (this was particularly noted in psychology, where there had been an increase in this type of assessment).

Where stretch and challenge is not firmly immersed within a thorough understanding of the subject through the use of focused questions or, for example, the development of specialist language and methods of enquiry; the higher-order skills may detract from rather than enhance subject knowledge, skills and understanding. In geography and psychology (and to some extent, English literature), the development of some methods of enquiry associated with the subject (such as close textual analysis of a text in English literature and research skills in geography and psychology) may be lost if teachers adopt the minimalist approaches.[27] Subject reviewers suggested this is possible by only covering the part of the specification expected to be assessed.

3.6 The A grade in terms of stretch and challenge and mastery of subject*

An analysis of stretch and challenge ratings across all candidate scripts was carried out against stretch and challenge ratings for A and A* candidates.[28] The results are summarised in Tables 4–8 and Figures 4–8. Please note that there are no tables and figures for English literature specifications 3 and 5 because some of the relevant data was not available.[29]

A2 UMS score was not available for several of the specifications, so it is not always clear why some A grade candidates with high UMS did not receive an A*. The emphasis on A2 scores in the awarding of the A* does mean that candidates cannot gain an A* without focusing on the (theoretically) more challenging A2.

Overall, it would appear that the A* is differentiating the most able candidates with a few exceptions, which are outlined in the discussion below.

3.6.1 French

One A grade candidate has a much higher stretch and challenge rating than one of the A* candidates. It is highly likely that the 90% A2 rule prevented the candidate from getting an A*, as their overall UMS score and mean stretch and challenge ratings are on a par with those of A* candidates. The stretch and challenge rating is acting as a distractor here. This is an example of a candidate who shows that stretch and challenge does not tie in perfectly with subject ability.

---

[27] In geography, for example, reviewers suggest this may manifest itself in formulaic approaches to essay writing as well as to cutting fieldwork beyond minimal preparation for the external assessment..

[28] Generally for 2008 we had six A grade scripts, while for 2010 we had three A grades and three A* grades (A* grades were awarded in 2010 but not 2008). It would have been possible to treat the top three 2008 scripts (by UMS score) as A*, but this would not necessarily have been strictly correct (it is not known if the candidates achieved at least 90% or more across the A2 units). We therefore look here at A and A* together.

[29] English literature 3: no UMS score available. English literature 5: no grade information available.

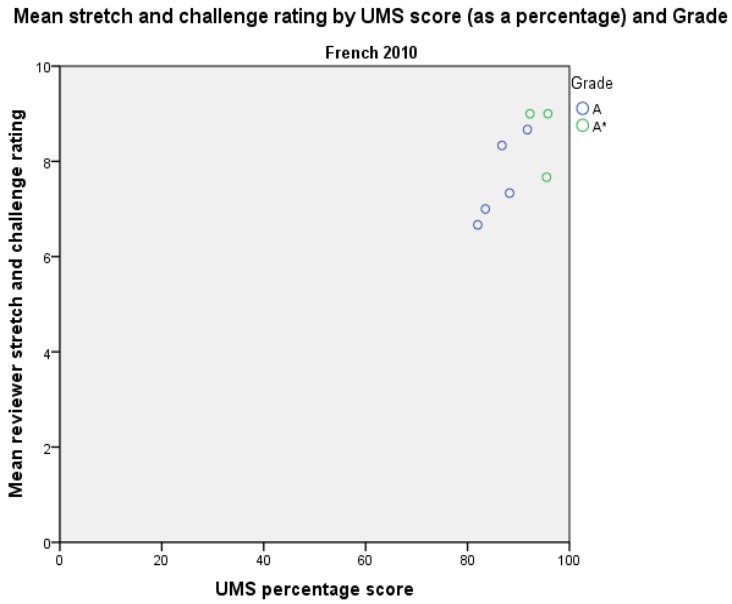

Figure 4: Mean stretch and challenge rating by UMS score (%) and grade: French

Table 4: Mean stretch and challenge rating by UMS score (%) and grade: French

	Year	Grade	UMS % score	Mean S&C rating
French	2010	A	0.88	7.33
French	2010	A	0.84	7.00
French	2010	A	0.82	6.67
French	2010	A	0.92	8.67
French	2010	A	0.87	8.33
French	2010	A*	0.96	9.00
French	2010	A*	0.96	7.67
French	2010	A*	0.92	9.00

3.6.2 Geography

One A grade candidate has a much higher stretch and challenge rating than one of the A* candidates (though this candidate had a very high UMS percentage score). It is most likely, as was the case with French, that the 90% A2 rule prevented the A candidate from getting an A*, as their overall UMS score and mean stretch and challenge ratings are on a par with those of A* candidates. The A* candidate script with the relatively low mean stretch and challenge rating (6.33) was discussed as part of the benchmarking exercise carried out at the second workshop. The script was ranked (reviewer mean average 11) and rated lower

than would be expected for an A* grade. The mean average had not been skewed by one reviewer giving substantially lower scores; the consensus was that the candidate offered descriptive rather than analytical analysis and showed limited conceptual understanding.

The analysis by subject reviewers suggests that in 2010 it was more difficult to differentiate student performance in some subjects, especially that of higher-attaining candidates. The reviewers reported that the geography candidates in 2010 across the grades showed a very formulaic and simplistic approach to essay writing; on a positive note, lower-attaining candidates could structure an essay. However, the lack of focus for many of the questions set in the assessment meant that it was difficult to differentiate between candidates in the upper quartile, and the mark schemes were not tight enough to enable this to occur.

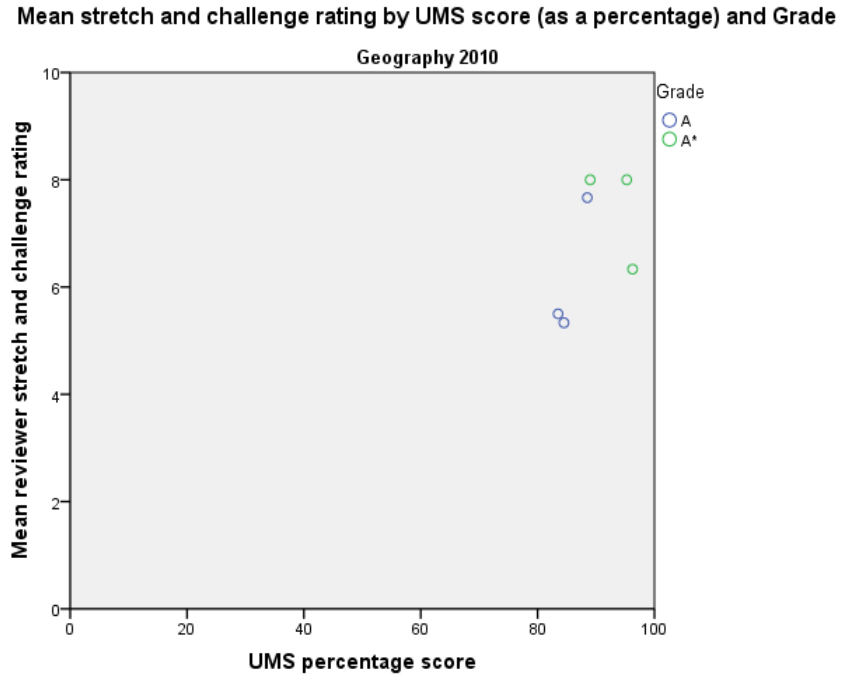

Figure 5: Mean stretch and challenge rating by UMS score (%) and grade: Geography

Table 5: Mean stretch and challenge rating by UMS score (%) and grade: Geography

	Year	Grade	UMS % score	A2 UMS % score	Mean S&C rating
Geography	2010	A	0.85		5.33
Geography	2010	A	0.84		5.50
Geography	2010	A	0.89		7.67
Geography	2010	A*	0.96		6.33
Geography	2010	A*	0.95		8.00
Geography	2010	A*	0.89		8.00

3.6.3 Media studies

One A grade candidate has a UMS score higher than the two A* candidates, but their mean stretch and challenge rating is lower. It looks as if the A* candidates have been differentiated well from the A candidates, although that same A candidate still has a high stretch and challenge rating, particularly when compared with the other A grade candidates. Without the A2 UMS score it is difficult to tell whether this candidate was close to receiving an A* grade and whether their A2 UMS score was perhaps the only reason they were not awarded one. Although, media studies candidates in 2010 could be differentiated, there was evidence that even the higher-attaining students were rarely going beyond descriptive recount in their evaluation for the A2 external paper. This was considered to be due to the invitation to 'describe' a process (or processes), which was not considered to be a suitable command word at this level in terms of intrinsic demand or discrimination.

Table 6: Mean stretch and challenge rating by UMS score (%) and grade: Media studies

	Year	Grade	UMS % score	A2 UMS % score	Mean S&C rating
Media studies	2010	A	0.87		3.00
Media studies	2010	A	0.84		6.00
Media studies	2010	A	0.90		8.25
Media studies	2010	A*	0.88		9.50
Media studies	2010	A*	0.87		9.00

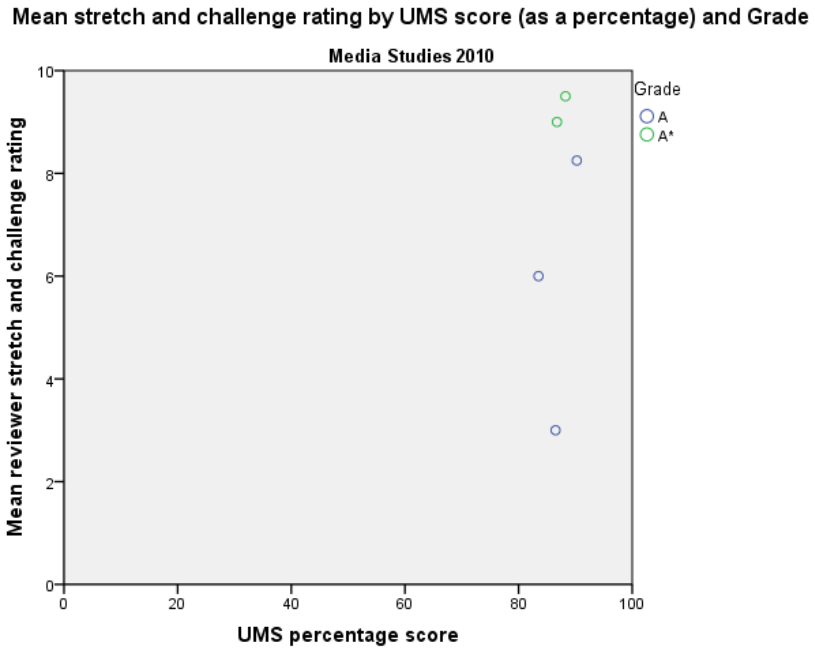

Figure 6: Mean stretch and challenge rating by UMS score (%) and grade: Media studies

3.6.4 Physics

One A grade candidate has very good mean stretch and challenge rating, and very good UMS score, both overall and at A2 (above 90% for both). Based on stretch and challenge and UMS score, this candidate looks like an A* candidate but was in fact only awarded an A.[30] The loss of one theory paper was judged to have resulted in fewer questions that might involve stretch and challenge for the really able, as the marks available were needed to provide discrimination among the current A level grades. This, therefore, was a change for the worse in the opinion of the subject reviewers in terms of assessing and establishing the real ability of all candidates. The increase in 'explanation' over mathematical calculation in physics meant there were fewer complex multi-stage calculations requiring some of the higher-order skills to stretch or reward the stronger mathematicians. It was noted, however, that all the candidates found the questions requiring explanation challenging.

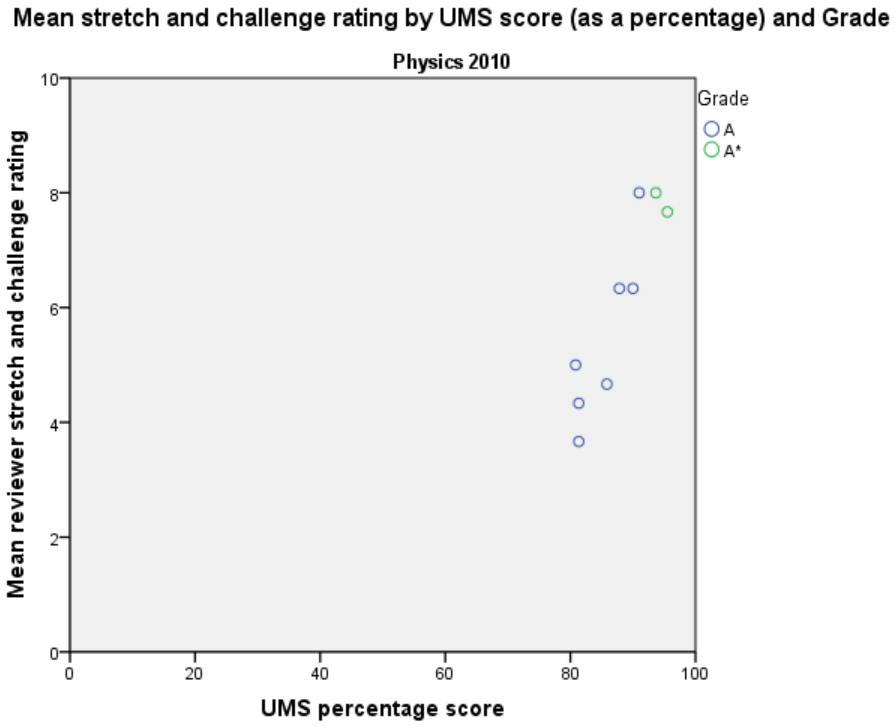

Figure 7: Mean stretch and challenge rating by UMS score (%) and grade: Physics

[30] Alternatively, this may be down to an error in the data sent to Ofqual by the awarding organisation, as other evidence strongly suggests that this candidate should have been awarded an A*.

Table 7: Mean stretch and challenge rating by UMS score (%) and grade: Physics

	Year	Grade	UMS % score	A2 UMS % score	Mean S&C rating
Physics	2010	A	0.81	0.72	5.00
Physics	2010	A	0.81	0.75	4.33
Physics	2010	A	0.81	0.78	3.67
Physics	2010	A	0.88	0.83	6.33
Physics	2010	A	0.86	0.85	4.67
Physics	2010	A	0.90	0.88	6.33
Physics	2010	A	0.91	0.93	8.00
Physics	2010	A*	0.94	0.95	8.00
Physics	2010	A*	0.96	0.98	7.67

3.6.5 Psychology

There appears to be some overlap between the A and A* candidates here. In particular, one candidate has a mean stretch and challenge rating and a UMS score higher than that of one of the A* candidates, but has been awarded an A apparently because their A2 UMS scores were not high enough. In this case, the mechanical calculation of the A* grade appears to have disadvantaged some candidates who may otherwise have been awarded an A*. The relatively low stretch and challenge rating reflects the subject reviewers' concerns that the 2010 papers offered less opportunity for stretch and challenge.

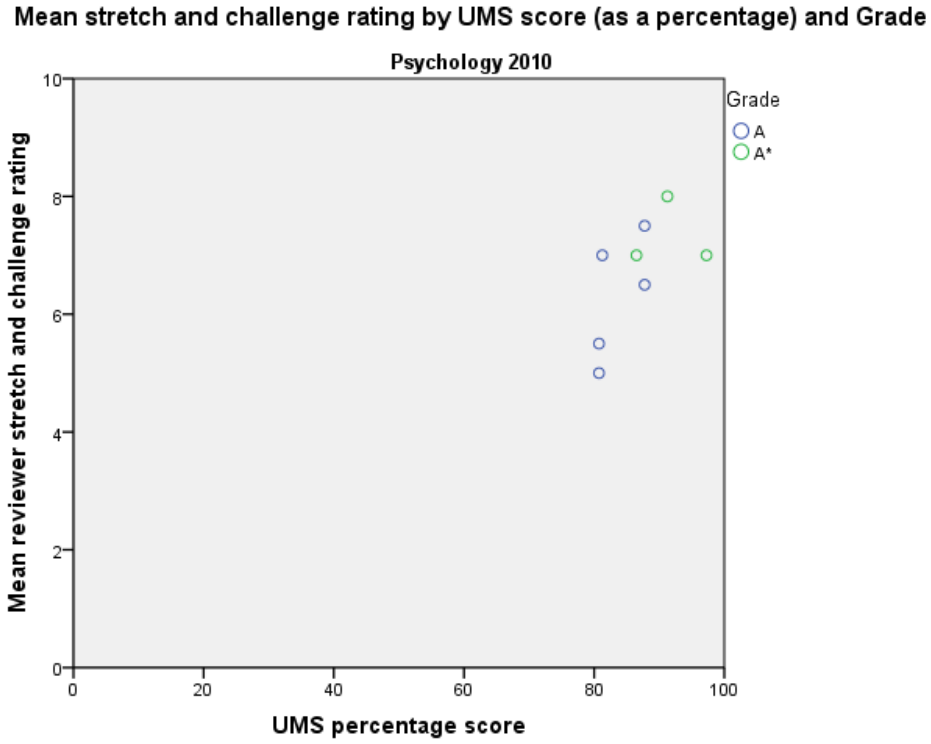

Figure 8: Mean stretch and challenge rating by UMS score (%) and grade: Psychology

Table 8: Mean stretch and challenge rating by UMS score (%) and grade: Psychology

	Year	Grade	UMS % score	A2 UMS % score	Mean S&C rating
Psychology	2010	A	0.81	0.75	5.00
Psychology	2010	A	0.81	0.78	5.50
Psychology	2010	A	0.81	0.88	7.00
Psychology	2010	A	0.88	0.83	6.50
Psychology	2010	A	0.88	0.85	7.50
Psychology	2010	A*	0.87	0.93	7.00
Psychology	2010	A*	0.91	0.96	8.00
Psychology	2010	A*	0.97	0.98	7.00

3.6.6 English literature (1)

One A* candidate has a mean stretch and challenge rating and UMS score that are not particularly high compared with A grade candidates, but the candidate managed to average above 90% in their A2 modules and so received an A* grade. Another candidate, despite getting a high stretch and challenge rating, did not average an A grade at A2.

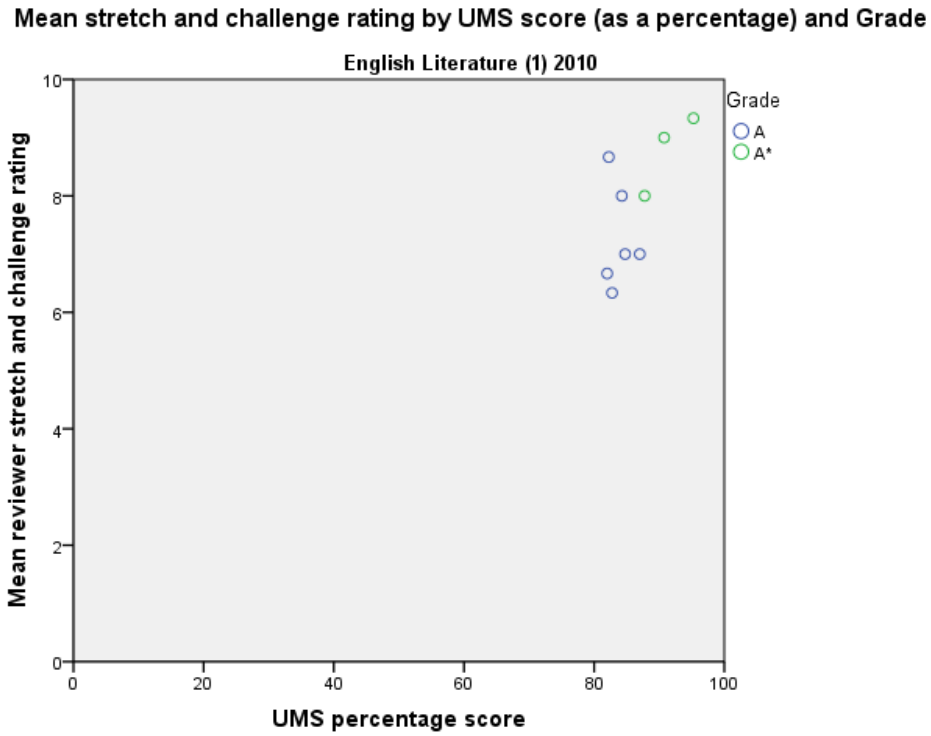

Figure 9 : Mean stretch and challenge rating by UMS score (%) and grade: English literature (1)

Table 9: Mean stretch and challenge rating by UMS score (%) and grade: English literature (1)

	Year	Grade	UMS % score	A2 UMS % score	Mean S&C rating
English lit (1)	2010	A	0.82	0.73	6.67
English lit (1)	2010	A	0.83	0.76	6.33
English lit (1)	2010	A	0.87	0.85	7.00
English lit (1)	2010	A	0.85	0.88	7.00
English lit (1)	2010	A	0.82	0.77	8.67
English lit (1)	2010	A	0.84	0.83	8.00
English lit (1)	2010	A*	0.88	0.93	8.00
English lit (1)	2010	A*	0.91	0.95	9.00
English lit (1)	2010	A*	0.95	0.98	9.33

3.6.7 English literature (2)

One A grade candidate has an extremely high mean stretch and challenge rating compared with even the A* candidates, and their UMS score is also high. Without the A2 UMS scores, however, it is difficult to say what prevented them from receiving the A* grade – they could have had a borderline A2 UMS score.

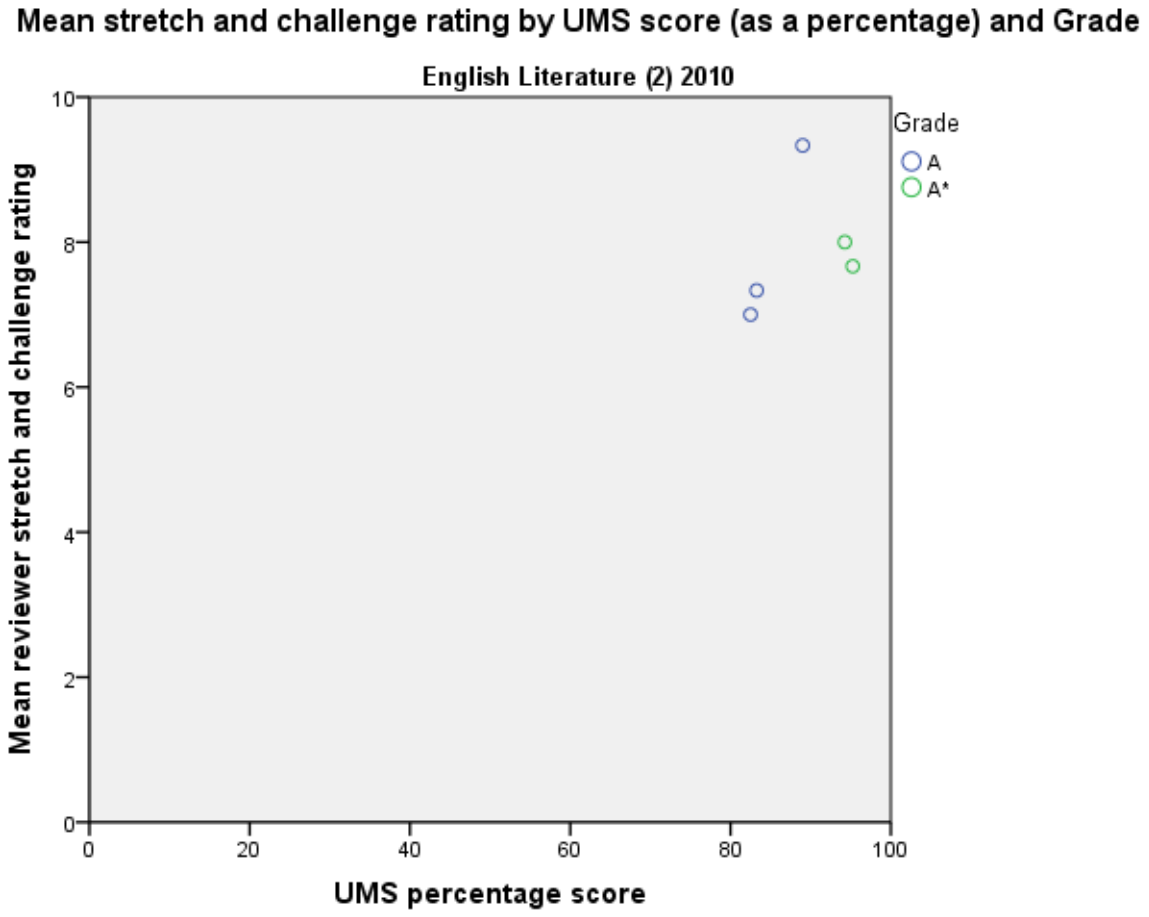

Figure 10: Mean stretch and challenge rating by UMS score (%) and grade: English literature (2)

Table 10: Mean stretch and challenge rating by UMS score (%) and grade: English literature (2)

	Year	Grade	UMS % score	A2 UMS % score	Mean S&C rating
English lit (2)	2010	A	0.83		7.33
English lit (2)	2010	A	0.83		7.00
English lit (2)	2010	A	0.89		9.33
English lit (2)	2010	A*	0.95		7.67
English lit (2)	2010	A*	0.94		8.00

3.6.8 English literature (4)

With this specification the A* candidates have good UMS and good stretch and challenge rating. Where the UMS score is good, the stretch and challenge usually is very high; if stretch and challenge rating is good, then UMS is usually very high. In general, it looks as if A* candidates are being differentiated well from the A candidates.

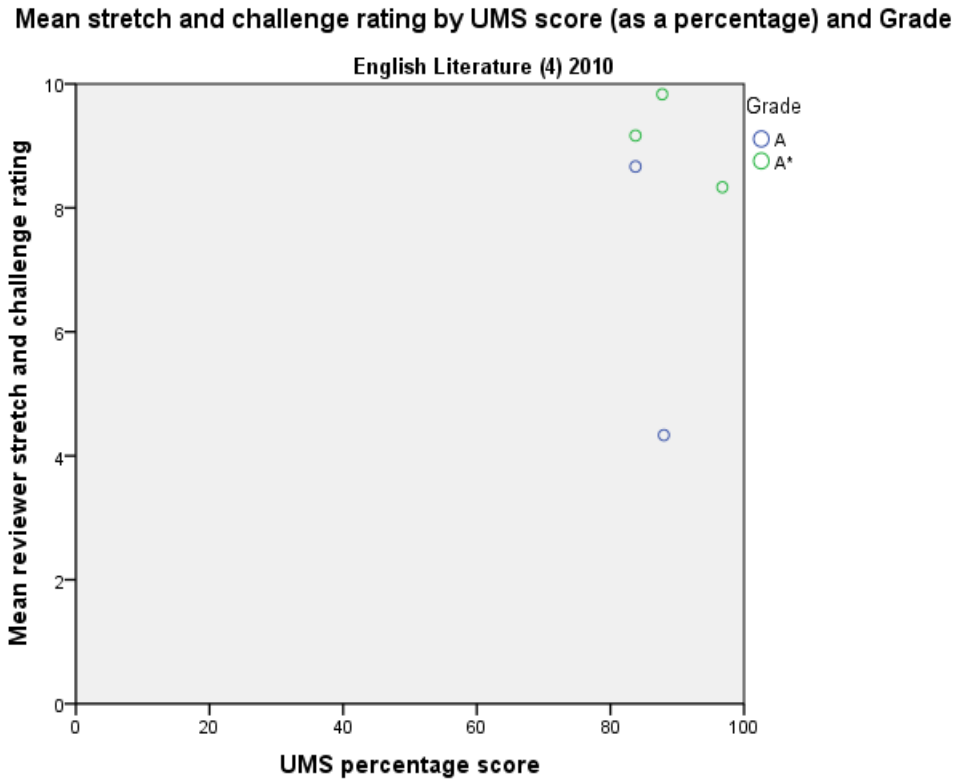

Figure 11: Mean stretch and challenge rating by UMS score (%) and grade: English literature (4)

Table 11: Mean stretch and challenge rating by UMS score (%) and grade: English literature (4)

	Year	Grade	UMS % score	A2 UMS % score	Mean S&C rating
English lit (4)	2010	A	0.88	0.85	4.33
English lit (4)	2010	A	0.84	0.88	8.67
English lit (4)	2010	A*	0.84	0.93	9.17
English lit (4)	2010	A*	0.88	0.95	9.83
English lit (4)	2010	A*	0.97	0.98	8.33

3.7 Evidence of development of skills between AS and A2

There has to be caution when considering the extent to which progression in the development of skills between AS and A2 across the two years reviewed can be analysed with confidence, given that scripts for AS and A2 were for matched candidates (on UMS scores) rather than a whole body of work for an individual candidate. However, some analysis of demand between AS and A2 across the two years has been conducted and therefore whether the demand at AS is likely to support progression.

The gap between AS and A2 is perceived to have **widened** in English literature. The reviewers judged that the AS assessment in 2010 exerted a considerably lower level of demand on candidates than the AS assessment in 2008. The reduction in demand at AS level in 2010 was seen to be attributable to less-challenging texts and an increase in short-answer and predictable questions in the assessments. At the same time, A2 units assessed in 2010 were judged to have shown increased stretch and challenge on 2008, generated by novel approaches to task setting and questioning.

The gap between AS and A2 is perceived to have **slightly widened** in psychology, too, because the 2010 AS examination was less demanding than that of 2008 and the 2008 AS examination was thus better preparation for A2 than the 2010 AS examination. The reason given for this was that the A2 question papers rely heavily on essay questions, and extended writing is significantly reduced in the 2010 AS examination. However, the 2010 AS examination was also judged to include some positive elements in comparison with that of 2008, including a wider variety of questions, albeit ones requiring only short answers, and an increased emphasis on assessing candidates' skills.

In French, the AS got slightly less demanding from 2008 to 2010, as the cultural element was moved to A2. Although the gap between AS and A2 might have **widened slightly**, the subject lead reviewer thought that there is now better, clear progression for students. In the lead reviewer's judgement, there is now a significant difference between the range, accuracy and complexity of the language skills developed over the two levels, and the cultural element is also better suited to A2 than AS.

The gap between AS and A2 is perceived to have got **narrower** in geography because, in 2010, AS assessments were more demanding, and A2 assessments less demanding, than in 2008. Where in 2008 there were only mini-essays in AS assessments, in 2010 candidates were required to write 25-mark essays. It was also noted, however, that there was no discernible difference in the quality of the essays at A2 between the two years, despite the increased emphasis on extended writing after 2008. On this basis, the lead reviewer concluded that the extended writing at AS is not developing candidates further in terms of stretch and challenge. The lower level of demand in A2 assessments in 2010 was judged to be in part due to the 'issues questions' not requiring candidates to apply their knowledge and understanding.

In physics, the gap between AS and A2 is also perceived to have got **narrower**, because AS exams in 2010 were slightly more challenging (and thus good preparation for A2) and the two most difficult elements of AS (2008) are now in the A2 exam in 2010. The slight increase in challenge in the AS exams in 2010 was seen to be due to the 6-mark-answer questions and the explanation questions, which require detailed use of scientific language and had

previously been included only at A2 assessments. It was judged that there was a better progression between AS and A2 in 2010, as the topics on kinetic theory and momentum, which weaker candidates found difficult at AS in Unit 2 in 2008, had moved to A2 in 2010.

In media studies, the AS in 2010 was more demanding than 2008 and the gap between AS and A2 had got **narrower**. In the reviewers' judgement, there is a disjunction in the progression in application of the knowledge of media production process and the skill to execute it between AS and A2 in 2010. Reviewers were, however, unable to explain this lack of progression. The reviewers raised the following questions: Are the production demands – in quantity terms – too great at A2? Does the move to moving image mean that the candidates do not see the relevance of the research/planning and evaluation work they have learned about at AS?

3.8 The extent and impact of variation observed within English literature specifications across awarding organisations

The inclusion of an English literature specification from each of the awarding organisations, although not directly linked to one of the research questions, makes it possible to consider the variation in specification within a subject.

As might be expected, given the variation possible across specifications, there was some difference in approach observed across the five awarding organisations. There were significant changes to the specifications in 2010, these are:

- the structure was reduced from six units to four
- coursework moved from 30% of the final assessment to 40%
- the assessment objectives were rewritten to place more emphasis on wider reading, comparison and the connections between texts.

The outcomes from these new specifications have not been identical. English specification (3) clearly now offers a more demanding assessment; specifications 1, 2 and 4 are providing opportunities for able candidates to fulfil their potential. There is no evidence of less stretch and challenge in 2010, but stretch and challenge is revealed in different ways – text choice, task selection and a confident approach by candidates to unseen materials. There is some evidence that there is now a reduced emphasis on close textual reading, and there is an anxiety that a reductionist approach has gone too far in some of the specifications, so that there are shortcut routes for teachers and candidates.

In English literature (2), the same range of skills was broadly evidenced in both 2008 and 2010 but the review indicated that, in some areas at least, the 2008 specification and examination papers seemed to give candidates more opportunity to display these skills in an appropriate way. In particular, this was true of the synoptic assessments, which focused on analysis and synthesis.

To some extent, difference in specifications in 2010 may be due to the opportunity the changes offered to address a range of issues within certain specifications. English literature (3), for example, was considered not to be challenging enough in 2008, so changes in 2010

brought it up to the expected level of demand. Conversely, one A2 unit in English literature (4) in 2008 was considered too challenging; the 2010 specification addressed this issue.

The findings suggest that there should not be any generalisation about *all* specifications for a subject based on the analysis of a single specification from one awarding organisation or, indeed, about all specifications for a subject from the same awarding organisation. This suggestion is also confirmed by subject reviewers' familiarity with specifications from other awarding organisations. There is no uniform approach to the interpretation of subject criteria by awarding organisations.

4 Summary and conclusions

4.1 The impact of reducing the number of units and the introduction of stretch and challenge

Research question 1

Have the change from six to four units and the introduction of stretch and challenge been effective in improving breadth of knowledge and understanding of the subject?

One of the main findings of this study relates to the introduction of stretch and challenge at the same time as the reduction in assessment (typically from six units to four). The net effect of introducing these two changes concurrently has not been entirely uniform across subjects, but all subjects have been affected by the changes, and often the reduction in assessment is seen to have worked against the introduction of stretch and challenge.

Several of the 2008 specifications were already considered by the subject reviewers to offer the stretch and challenge required by the changes in 2010. There was evidence to show that the change of specification in 2010 has enabled 2008 specifications that were either too demanding or not demanding enough to be brought into line with other specifications. The issue for some specifications that were already considered demanding is that the further introduction of extended writing questions dislocated stretch and challenge from subject knowledge and skills.

In English literature, geography, physics and psychology there was the loss of a unit focused on synoptic assessment. Most reviewers saw this as a retrograde step because, although in theory for 2010 synopticity was intended to be embedded across A2, there were often issues identified in practice.

- Reviewers reported that the removal of the synoptic unit meant that physics candidates in 2010 had less opportunity to show an overall grasp of the course and to show their ability to connect different topics. This made synoptic assessment across the course more problematic.
- In geography the range of unit options was thought to lead to fewer opportunities to make links across the different elements of the course. This made synoptic assessment across the course more problematic. Although the essay-style questioning attempted to promote synoptic assessment in 2010 (creating connections), it could be argued that this was at the expense of the stretch and challenge that relates directly to the subject as a discipline (use of specialist language and methods of enquiry).
- In psychology the 2010 paper contains more structured questions and more short answer questions. As a result reviewers felt that there were fewer opportunities for able candidates to demonstrate their skills and reduced opportunities for candidates to make links between different parts of the specification.

- Across both years in English literature there were opportunities for synoptic assessment implicit within the skills of the discipline (e.g. comparing and contrasting texts and thematic approaches to textual analysis).

Research question 2

Is there any evidence of stretch and challenge in the responses across the grade levels?

The stretch and challenge indicators, while generic higher-order skills, contain elements that are related to a subject or discipline, such as the specialist language or the different methods of enquiry used within the academic and empirical field of study. The positive correlation between levels of stretch and challenge and grade may suggest that stretch and challenge and breadth and understanding of the subject are working together. The qualitative data, however, suggests that for the majority of subjects – with media studies and, to some extent, French the exceptions – some aspects of stretch and challenge prioritised are not those considered by the subject reviewers as important aspects of the subject as a discipline. For example:

- English literature (all specifications reviewed): the greater emphasis in 2010 on identification of related concepts and making comparisons was thought to be at the expense of evidencing close textual analysis of a single text – a method of enquiry fundamental to the discipline of English literature.
- Geography and psychology: reviewers observed that stretch and challenge in 2010 included a limited reference to the research skills required for these disciplines. The loss of coursework meant there was less evidence of stretch and challenge in terms of the methods of enquiry specific to the discipline. Geography reviewers also considered there to be less emphasis on the scientific nature of geography.
- Physics: the increased emphasis on explanation was considered to be at the expense of stretch and challenge in terms of the mathematical skills (specialist language and methods of enquiry) that support the work of a physicist.
- French: the increase in language skills required was at the expense of an in-depth, independent study of a text from the culture, which, it could be argued, builds skills required for wider understanding and meaning within the cultural context.
- There was some evidence across candidates' work assessed by a piece of extended writing (e.g. geography) that stretch and challenge could reward the ability to structure an essay (construction of an argument) at the expense of demonstrating knowledge and understanding of the subject.

The analysis of the scripts suggested that there is a strong positive relationship between A level UMS score and the level of stretch and challenge evidenced within the candidate work, which supports the view that stretch and challenge is being seen at all grades. It was relatively unusual for the stretch and challenge rating of even the lowest-ranked candidates to be below 3 on a 10-point scale used. It clearly suggests that stretch and challenge is present in the work of all candidates, while the extent to which it is displayed depends strongly on the overall level of subject knowledge and understanding.

There is not sufficient evidence to suggest that the introduction of stretch and challenge has had a widespread positive impact on breadth of knowledge and understanding of a subject, with the exception of media studies and, to some extent, English literature and French. Although stretch and challenge is evidenced in the work of higher-attaining candidates and mark schemes often award this, some aspects of stretch and challenge are prioritised above others to the detriment, in the view of the subject reviewers, of the study and understanding of the subject as a discipline.

Research question 3

How do candidates progress between AS and A2?

The pattern of progression between AS and A2 is uneven across the subjects. The gap between AS and A2 is perceived to have **widened** (to varying degrees) in English literature, psychology and French. The gap between AS and A2 is perceived to have got **narrower** (again, to varying degrees) in geography, physics and media studies.

These differences are the overall result of structural changes to the qualifications creating more or less demand at AS and/or A2. Such changes include the moving of a more challenging module to A2, a change in the type of examination questions (e.g. more or fewer essay-style questions or short-answer questions, a change in command words used), mode of assessment (e.g. removal of coursework), a change of emphasis of particular skills for a subject (e.g. more explanation and less mathematics in physics), or the removal of a synoptic unit.

These findings can give only a limited indication of possible progression between AS and A2, given that scripts for AS and A2 were for matched candidates (on A2 UMS scores) rather than a whole body of work for an individual candidate.

*4.2 Stretch and challenge evidenced at A**

Research question 4

Is there evidence that candidates can achieve A* without doing the stretch and challenge elements?

There are two aspects to this question. First, in subjects such as English literature, media studies and, to an extent, geography, the nature of the assessment consists almost entirely of extended writing. Here, as the banded mark schemes make clear, the stretch and challenge elements are almost impossible to disentangle from those specific to subject knowledge, and it would be almost impossible to score the kind of consistently high marks without showing considerable evidence of higher-order skills. In French, similarly, the nature of linguistic development means that it is essentially synoptic: again, it would be impossible to gain very high marks without, for example, the ability to synthesise. In subjects such as physics and, to an extent, psychology, which use a much greater proportion of short-answer questions focused on specific parts of the subject content, the precise theoretical answer to this question depends on exactly how much credit is given to questions meeting stretch and challenge requirements and where the raw grade boundaries are set. The question can be addressed in terms of the reviewers' ratings, which for stretch and challenge correlated very highly with the candidates' UMS scores. Moreover, the ratings for the A* candidates were

very high, typically 8, 9 or 10 on the 10-point scale. Thus, the evidence from this exercise is clear that most candidates did not achieve an A* without achieving considerable success with the stretch and challenge elements.

It should, however, be noted that the type of assessment is sometimes at odds with what some reviewers consider the nature of the subject. The higher-levels of stretch and challenge available in extended writing assessments in geography, for example, do not *necessarily* reward the best geographers unless they can write essays. These higher-order skills are prioritised above fieldwork skills in A level geography. Any findings therefore need to be considered within the wider context of decision making in relation to the skills, knowledge and understanding prioritised for a subject at this level of study.

Research question 5

Has the introduction of A* made it possible to differentiate the most able candidates?

It has already been noted that the correlations between UMS scores, reviewers' rankings and reviewers' ratings were very high, suggesting that there was a high degree of consensus about the most able candidates, whether viewed in terms of general subject ability or higher-order skills. With very few exceptions reviewers ranked the A* candidates right at the top of the 2010 candidates (typically 8, 9 or 10 on the 10-point scale), irrespective of how they compared with the 2008 candidates. However there were individual anomalies in most subjects.

- One French A grade candidate has a much higher stretch and challenge rating than one of the A* candidates.
- In geography, the data shows that one A grade candidate has a much higher stretch and challenge rating than one of the A* candidates.
- One media studies A grade candidate has a UMS score higher than the two A* candidates.
- One A grade physics candidate has very good mean stretch and challenge rating, and very good UMS score.
- One psychology candidate has a mean stretch and challenge rating and a UMS score higher than that of one of the A* candidates.

In most of the cases listed above further analysis made it seem likely[31] that the 90% A2 rule prevented the anomalous A grade candidate(s) candidates from getting an A*, as their overall UMS score and mean stretch and challenge ratings are on a par with those of A* candidates. The available evidence therefore suggests that the A* did generally differentiate the most able candidates, with the caveat that the 90% A2 rule appears to have impacted on some candidates who may otherwise have been awarded an A*.

[31] The data was not available in all subjects to support full analysis.

Research question 6

What does A* performance look like?

Findings suggest that the A* generally rewards consistently high performance and evidences both higher-order skills and understanding of the subject at A2. Further data would be required to provide a more robust explanation for some anomalies – for example, where students appear to warrant an A* grade based on the quantitative and qualitative data available but have not been awarded the grade, or where a very few students appear to have an A* without a high rating for stretch and challenge. Given the small number of reviewers and scripts, however, it is not possible to judge whether anomalies are likely to be systemic or not.

With the necessary caveats, A* candidate performance could be described as displaying many of the higher-order skills effectively within the context of the subject, as characterised by the stretch and challenge indicators:

- creating connections: identifying related concepts and making comparisons, generalisation, transfer and recontextualisation
- constructing an argument
- using explanation, application and synthesis of ideas rather than just recall of facts
- using strategies for investigation and problem solving
- using, and understanding, specialist language and methods of enquiry.

4.3 Concluding remarks

The research reported here focuses on evidence of any difference, or similarities, between the 2008 and 2010 specifications and students' performance in the examinations, with particular reference to stretch and challenge and the introduction of the A* grade.

Several of the 2008 specifications studied were already considered by the subject reviewers to offer the stretch and challenge required by the changes in 2010. One of the key findings of this study is that the assumption that an increase in questions requiring extended writing will identify greater stretch and challenge and breadth of subject knowledge and skills has not always been realised in candidate performance.

A second key finding has been in relation to the introduction of stretch and challenge at the same time as the reduction of units (typically from six units to four): reviewers felt that the elements of stretch and challenge have often been compromised by the reduction in assessment. Where stretch and challenge is not firmly immersed within a thorough understanding of the subject through the use of focused questions or, for example, the development of specialist language and methods of enquiry, then the higher-order skills may detract from rather than enhance subject knowledge, skills and understanding.

Synoptic assessment was also frequently mentioned by reviewers. Awarding organisations had been asked to test understanding and connectivity through synoptic questions, but in parallel there was the loss of a synoptic unit in a number of specifications. Reviewers often felt that the loss of the unit led to less overall opportunity for candidates in 2010 to show an

overall grasp of the course and fewer opportunities to make links across the different elements of the course.

The inclusion of an English literature specification from each of the awarding organisations made it possible to consider the variation in specification within a subject. The findings suggest that there should not be any generalisation about *all* specifications for a subject based on the analysis of a single specification from one awarding organisation or, indeed, about all specifications for a subject from the same awarding organisation. This suggestion is also confirmed by subject reviewers' familiarity with specifications from other awarding organisations. There is no uniform approach to the interpretation of subject criteria by awarding organisations. However this research has identified key issues which may be seen, to a greater or lesser extent, across specifications and subjects.

We wish to make our publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by the Office of Qualifications and Examinations Regulation in 2012

This publication is also available on our website at www.ofqual.gov.uk .

Any enquiries regarding this publication should be sent to us at:Office of Qualifications and Examinations Regulation

Spring Place
Coventry Business Park
Herald Avenue
Coventry CV5 6UB

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

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