



**STUDENTS' VIEWS OF STRETCH AND CHALLENGE IN
A-LEVEL EXAMINATIONS**

**Jo-Anne Baird, Suzanne Chamberlain, Michelle Meadows, Lucy Royal-Dawson,
& Rachel Taylor**

March 2009

Contents

1	Executive Summary	3
2	Introduction	4
2.1	Stretch and challenge policy	5
2.2	Policy implementation guidance	5
2.3	Can examiners write questions at a particular level of demand?	6
2.4	Motivational effects of external assessments	8
2.4.1	The motivational impact of formulaic, narrow assessment	9
2.4.2	The motivational impact of challenge and extended writing	11
2.5	Assessing gifted and talented students	12
2.6	Students' and teachers' views about A-levels	14
2.7	The current study	15
3	Methodology	15
3.1	Sample	15
3.2	Materials	16
3.3	Procedure	17
3.4	Analysis	17
4	Findings	18
4.1	Students' reactions to the stretch and challenge specimen papers	18
4.1.1	Layout	19
4.1.2	Question type and content coverage	19
4.2	Examination preparation	22
4.2.3	Aspects of difficulty	25
4.3	Motivation and stress among high achievers	26
4.4	Challenge in examinations	27
5	Discussion	28
6	Conclusion	30
7	Acknowledgements	30
8	Disclaimer	31
9	References	31
Appendix A	Recruitment e-mail	35
Appendix B	Interview schedule notes	36
Appendix C	Consent form	39
Appendix D	Information sheet	40
Appendix E	Pedagogical approaches to exam preparation	41

Tables and Figures

Figure 1	Estimates of success rates on items and the actual outcomes for A-level Economics questions (from Meyer, 2003)	7
Table 1	Subject under investigation and university attended	16
Table 2	Schools and colleges of origin for the students in the sample	16
Table 3	Ratings of specimen papers compared to summer 2008 A-level question papers	19

Stretch and challenge evaluation

1 Executive Summary

Claims that A-levels do not stretch the brightest students¹ have dogged the education system in England in recent years. Part of the Government response to this has been to introduce a policy of 'stretch and challenge', which features different question styles in new A-level examination papers. The questions are not intended to produce more difficult A-level examinations overall, but to foster more engagement, broader thinking and higher order skills in students. Teaching-to-the-test, narrow and formulaic approaches are intended to be circumvented by the new-style examinations. QCA produced guidance for the Awarding Bodies on how to implement stretch and challenge in question paper design.

This research focussed upon that implementation, in A-level psychology and biology specimen question papers from all three English Awarding Bodies. Previous research literature did not provide a firm evidential basis for examiners to predict how demanding a question is, nor is there evidence to suggest that examiners are good at judging question difficulty. Worryingly, previous research indicated that students could be de-motivated by challenging examination questions.

Thirty-nine students who had sat their A-level examinations in summer 2008 were asked about their reflections of A-level study, their views of last summer's question papers and were asked to contrast the new-style question papers with the old ones. Overwhelmingly, students reported being drilled in examination techniques and recognised that, although this had merits, it had interfered with the breadth and depth of their studies. Some teachers had demonstrated considerable creativity in this drilling, as innovative teaching approaches were also reported. The picture of student experiences was similar to that reported in earlier research on the first cohorts of students to undergo the *Curriculum 2000* A-levels. Comments like the following typified students' approaches to learning,

"You have to know exactly what [the examiners] want to hear from you." Biology student

'After doing A-levels you sort of realise that A-levels aren't about how good you are at the subject. It's how good you are at learning the mark scheme.'

Psychology student

By allowing students access to the marking scheme, Awarding Bodies have spread what Bourdieu (1986) termed cultural capital. This form of capital is formed of knowledge and skills that are useful for advancing an individual's status in society. One question raised in this research is whether it is cultural capital of the right kind that is currently being passed on to students. This philosophical issue goes to the heart of what education is for and what kinds of knowledge we value as a society.

Although students did not report an obvious change in the demand of the new-style examinations, they did spot features that are congruent with the stretch and challenge policy. For example, applied questions were unpredictable and students considered that they would have to think flexibly about applying their knowledge in the examinations. Shorter questions, they thought, would mean that they

would have to study the subject more broadly. Also, the question papers appeared to the students to have a broader content coverage of the syllabus and to be wordier. This research involved student opinion and it is possible that performances on the examinations and approaches to study will not be in line with students' initial reactions in hour-long interviews.

Despite some encouraging signs that the new psychology and biology examinations might achieve their policy objectives, it will be important to monitor the impact of the new examinations upon teaching and learning following their introduction. After all, it is only once teachers have experienced the new examinations and had the time to adapt their approaches that the backwash effects might be found. The possibility that undesirable styles of teaching and learning were caused by something else in the education system that cannot be influenced by changing examination question styles remains.

2 Introduction

In recent years, some commentators have argued that A-levels do not stretch the most able students, as the following quotation from Richard Cairns, Magdalen College School in Oxfordshire illustrates.

"Pupils are thinking they need to do more and more in order to differentiate themselves from others so they are piling on more subjects rather than trying to stretch themselves by doing something different and challenging." (BBC, 2005)

At times, this allegation has been levelled at the nature of the assessments – either their modular format (eg Paris, 2007: 'Winchester drops too easy A-levels', *The Telegraph*), or the questions themselves (eg Paton, 2009: 'Bright pupils agonising over simple questions says Eton Head', *The Telegraph*). To tackle such concerns, the Government introduced a policy that has become known as 'stretch and challenge' in the *Education and Skills White Paper* (DfES, 2005). New A-level examinations were introduced, for first teaching in September 2009, which incorporated questions designed to induce more stretch and challenge in students' examination experiences. This research investigated students' perceptions of the new A-levels and their experiences more broadly of studying for, and sitting, A-level examinations.

In this introductory section, we outline the nature of the policy and guidance on the implementation of the policy. We then turn to what we know from the research literature on designing demanding examination questions. Motivation of students to study subjects in depth appears to be an underlying assumption of the policy, so we investigated whether the research literature indicates particular relationships between summative assessments and motivation. As the policy was designed to stretch the most able students, we investigated whether there was anything instructive in the research literature on assessing gifted students. To contextualise the research further, we also looked to find out what was known about students' reactions to A-level examinations. First, let us outline the stretch and challenge policy.

¹ Defined as those at least meet the requirements for a grade A for the purposes of this research.

2.1 Stretch and challenge policy

Paragraph 8.15 of the White Paper introduces the issue as follows,

“First, we want more stretch within A-levels. Because we make it a priority to preserve A-level as a qualification, with consistent standards over time, we will take a slightly different route to that proposed by the Working Group. We will seek the introduction of a new section in A-level papers covering AEA material. We will ask QCA to consider the best means of doing this across all A-levels, so that increased stretch and challenge are available to all students in all types of institution, and scholarship can flourish.”
DfES, 2005, p59-60

Preceding paragraph 8.15 in the White Paper, there was a reference to the lack of discrimination in A-level grading and the ensuing problems for selection to Higher Education. But the stretch and challenge policy was not designed to solve that problem, as ‘greater differentiation’ was considered elsewhere in the White Paper and, ultimately, the decision to introduce an A* grade was designed to deal with that problem. Ruth Kelly (2006), the then Secretary of State for Education and Skills, wrote that her primary purpose in introducing the stretch and challenge policy was to “ensure that our brightest A-level students ... develop broader skills and knowledge than those currently required by A-level”. But the difficulty could not impact upon the grading of the examinations, as the consistency of A-level standards over time was expected. Therefore, the policy must have been intended to increase the experience of difficulty for the students – stretching them, but not necessarily rewarding them differently. This distinction relates to the discussion on the difference between demand and difficulty, which is outlined below.

It was intended that the stretch and challenge initiative would partly deal with criticisms that A-levels were too easy for the brightest students. However, there was reason to question whether A-level students really did experience the examinations as easy. Few students scored very high raw marks on the *Curriculum 2000* examinations. In some subjects, a small proportion of students were awarded full marks, in terms of module scores (uniform mark scores), but this was a function of the conversion of the original examiners’ marks into a credit system that made modular assessment structures feasible, ironing out the vagaries of changes in difficulty of the examinations. On the raw mark scale, students did not typically score highly and therefore their experiences of the examinations must have been challenging already. For example, in AQA, across 501 summer 2007 A-level assessments, the average proportion of marks that had to be scored for a grade A to be awarded was 73 *per cent*, leaving over a quarter of the raw mark scale to challenge the brightest students. For these reasons, the current research investigated students’ views of the challenge of the *Curriculum 2000* examinations, as well as their views of the new stretch and challenge questions. The research tested the need for the new policy, as well as how well students perceived it to have been implemented.

2.2 Policy implementation guidance

After a great deal of policy debate between various stakeholders, including an exchange between QCA and the government (Curnock-Cook, 2005; Kelly, 2006) and piloting, it was decided that different question styles would be incorporated into (A2) question papers in the second year of A-levels, to foster more extended writing, synthesis of study and less formulaic responding (see http://www.qca.org.uk/qca_10331.aspx). QCA advice indicated that the stretch and challenge questions should,

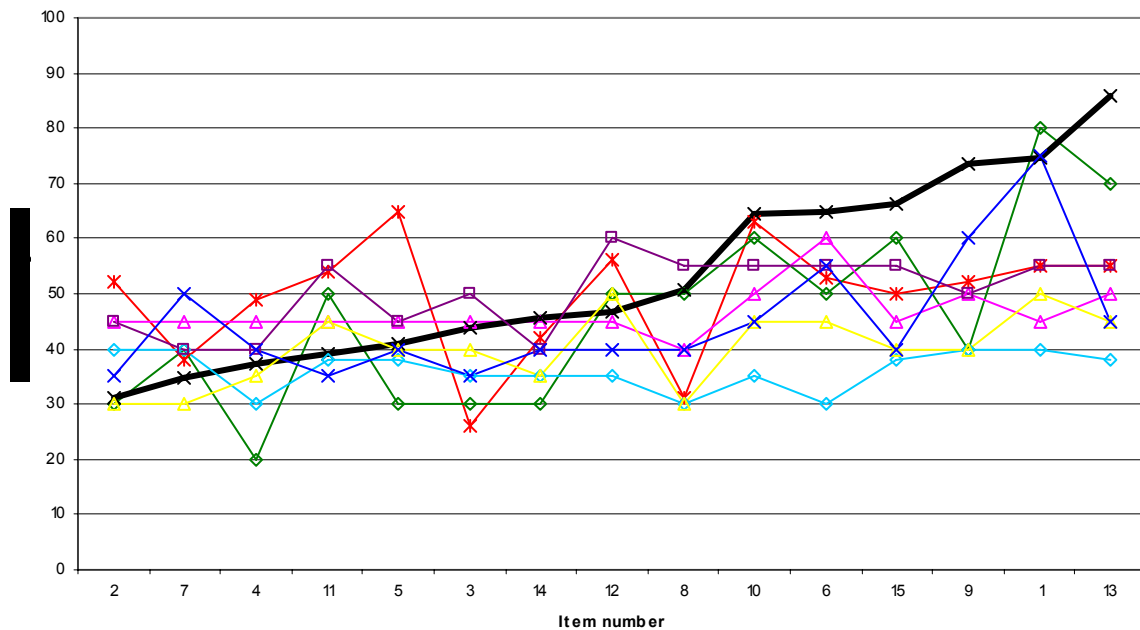
-
- 1) Use a **variety of stems in questions** to avoid a formulaic approach through the use of such words as: analyse, evaluate, compare, discuss
 - 2) **Avoid assessments being too atomistic**, connections between areas of content being used where possible and appropriate – there will not be a requirement to have overall coverage of all content area or dependence between the different sections which would mean failure on the first part of the question dooms the performance on following parts
 - 3) Have some requirement for **extended writing** in all subjects except where this is clearly inappropriate eg mathematics
 - 4) Use a wider **range of question types** to address different skills i.e. not just short answer/structured questions. The use of open answer questions, questions based on case studies and stimulus materials are encouraged
 - 5) Use improved **synoptic** assessment, asking candidates to bring to bear knowledge and the other prescribed skills in answering questions rather than simply demonstrating a range of content coverage.

QCA's policy guidance indicated that stretch and challenge was related not only to demand, but to the predictability and atomisation of assessment. So the first challenge for examiners in implementing this guidance is to write challenging questions in the new A-level assessments. The next section investigated what was known from the research literature about how to write questions targeted at a particular level of demand.

2.3 Can examiners write questions at a particular level of demand?

To be able to write questions at a particular level of demand, examiners need to be able to judge the demand of questions. Whilst it is often assumed that this comes naturally to teachers, let alone examiners, research evidence indicates that neither are good judges of the difficulty of items (eg Impara and Plake, 1998; Good and Cresswell, 1988; Meyer, 2003). Typical findings on this topic include the results from Wolf *et al.*'s (1995) study, in which a panel of experts were asked to rate the difficulty of 30 mathematics questions. These experts' ratings were then correlated with empirical values of difficulty of the items calculated on the basis of 301 students' performances on the test. The overall correlation was statistically significant, but moderate (approximately 0.4: Table 1, p347), indicating that these experts were not able to judge well the order of difficulty of the items for students. Good and Cresswell's (1988) research was more encouraging because their examiners were at least able to judge which question papers were more, and which were less, difficult. The magnitude of the differences in difficulty was not judged well in Good and Cresswell's studies, however. Figure 1 shows data from Meyer's (2003) study of A-level Economics questions. In her study, A-level examiners were asked to judge the likely success rates on each multiple-choice question for students who would be awarded a grade E on the examination. This judgment task formed part of the Angoff process of standard-setting (see Chapter 6, Cizek and Bunch, 2007). The thick line on Figure 1 shows the empirical outcomes for students on the examination and each of the other lines represent estimates of the success rates from five senior examiners. Correlations were moderate and, as can be seen, there were discrepancies in the empirical and estimated success rates of students for these questions. Pollitt's work helps elucidate why it is difficult for examiners to judge the demand of questions and distinguishes between the concepts of demand and difficulty.

Figure 1 Estimates of success rates on items and the actual outcomes for A-level Economics questions (from Meyer, 2003)



Writing questions that stretch and challenge the most able candidates depends on being able to distinguish between, and control, sources of demand and difficulty (Pollitt, Entwistle, Hutchinson and de Luca, 1985). There is also the question of the separation of the definition of ‘difficulty’ and ‘demand’ of questions. The two concepts are clearly closely related and Pollitt *et al.* (2007b, p.193) suggest that in the case of some aspects of question design ‘more demand directly causes more difficulty’. Demand is argued to be a subjective and qualitative property of a question (Pollitt *et al.*, 2007b), and is essentially the height of the hurdle that the examiners set when they define the syllabus and question papers. Difficulty is how well the students are rewarded for their efforts and tends to be measured quantitatively, for example as an average score. To illustrate the difference between the two concepts, consider the case in which a school does not teach a particular topic – for example, Pythagoras theorem. Students who encountered that topic in the examination would find it very difficult, but this does not mean it was a demanding topic – with teaching, they might have found it easy. Difficulty is the quantitatively measured performance of students, whereas demand is the cognitive load of a topic. Student preparation, in the form of tuition or other study, separates demand and difficulty (see Figure 2). However, demand and difficulty are often not distinct in students’ experiences. We can only speculate about what causes certain topics to be challenging for individuals, but for groups of students, it might be possible to distinguish between the two concepts.

Figure 2 The relationship between demand and difficulty

		Demand	
		Low	High
Difficulty	Low	Easy topic	Lack of preparation
	High	Lots of preparation	Hard topic

Assessing demand is a complex process because a candidate's ability to answer a question is not solely a matter of his/her familiarity with the content, nor of the perceived demand of the question (Ahmed and Pollitt, 2001). Demands are not always intended, nor integral to the subject matter being assessed. Pollitt *et al.* (2007b) suggested that issues of question demand should underpin the construction of questions, papers and mark schemes. Considering demand at this early stage facilitates the creation of assessments with appropriate standards. There are several facets of a question that can be adapted in order to increase demand, especially in the case of stretch and challenge. These include the level of abstractness or generalisation; the nature of the task (whether it is a low level demand descriptive task or a higher demand evaluative task); and the extent of guidance provided to candidates in terms of selecting appropriate material to include in an answer and devising a strategy for their response (Pollitt *et al.*, 1985). Examiner and researcher reviews of GCSE examination questions by Murphy *et al.* (2007) and Pollitt *et al.* (2007a) found numerous instances of illegitimate sources of difficulty - facets of the questions that unintentionally hindered or benefited candidates in demonstrating what they knew. These included overly complicated or inaccessible language, restrictive mark schemes, and the inclusion of diagrams or pictures that seemed to serve no purpose other than to potentially confuse candidates (Pollitt and Ahmed, 2000; Pollitt *et al.*, 1985).

Dhillon (2003) argued that the process of manipulating question demand depended upon the application of the tacit knowledge or 'intuition' of the question writer because it was an inexact science. While this may be the case for many question writers, the literature offers writers some advice regarding the identification and manipulation of question demand. Pollitt *et al.* (2007b) list some of the facets of an assessment that contribute towards demand. These include:

- The amount of time relative to the amount of work required, with more work in less time being more demanding than being permitted more time for less work;
- The level of reading difficulty; and
- The application of working - or long-term memory. The extent of scaffolding or stimulus material that a question writer provides can increase or decrease the reliance on memory (Pollitt *et al.*, 1985; Pollitt *et al.*, 2007b) and hence the demand.

Despite the foregoing research and the widespread nature of examining, knowledge in this area is not at a stage where accurate predictions can be made about the likely difficulty of questions. Nonetheless, tentative recommendations have been made in the literature about which factors examiners should attempt to take into account (Pollitt *et al.*, 2007b).

2.4 Motivational effects of external assessments

Lurking under the stretch and challenge policy is the assumption that examinations **can** stretch and challenge students, motivating students to study more broadly, in more depth and with more

enthusiasm. Let us look at the literature on the effects of summative assessments upon students' motivation.

Perhaps the first assumption that needs to be investigated is whether examinations *ever* motivate students. In Harlen and Deakin-Crick's (2003) review paper, several sources are cited which lead the authors to conclude that students are de-motivated by summative assessment. When feedback is related to the individual (ego-involved) and when it is in the form of grades, marks or scores, it is particularly negative. Feedback relating to the task and tailored to the student's performances, which is not norm-referenced or presented in any competitive manner is seen as more positive (Black *et al*, 2003): that is, formative assessment.

Reay and Wiliam (1999) explored how perceptions of Key Stage 2 tests contributed to children's understanding of themselves as learners. There was evidence that even at ages 10 and 11, pupils had clear perceptions of how external assessment impacted upon the curriculum they were taught. For example, the pupils reported repeating maths, language and sciences 'over and over' again as preparation for the tests, and doing multiple practice tests, meaning they had less opportunity to study more 'interesting' subjects compared to in the previous term. Further, students' identities as learners were affected by the labelling that they perceived to be attached to grades. For low attaining students, this was a very negative experience, as indicated by the title of Reay and Wiliam's paper, *I'll be a nothing*. Suffice it to say that there is a predominant view in the UK research literature that summative assessments are de-motivating, although there is more debate on this issue in the international literature (see Popham, 1987; Airasian, 1988; Chapman and Synder, 2000).

An association has been reported between motivation and performance in examinations (for a review, see Sundre *et al.*, 2004, p7). Normally, we would interpret this as meaning that motivation caused performances to some extent, but the stretch and challenge policy raises a further possibility: that the test partly causes the motivation, which in turn partly causes the performance. This is known in the literature as the washback effect (Messick, 1996) – the impact of assessments upon teaching and learning. Often this is associated with negative consequences, such as restriction of the taught curriculum (Harlen and Deakin-Crick, 2003). In effect, the stretch and challenge policy is an attempt to use the washback effect of assessment in a positive manner. Again, this is not new - other authors have noted the purposive use of washback in UK Government assessment policies (eg James, 2000).

Sundre *et al.* (2004, p9) reported a range of studies showing low motivation in testing situations where the results were of little consequence to the students. Thus, high-stakes testing itself can impact positively upon student motivation, but the stretch and challenge policy goes further than this, by assuming that *specific design features* of high-stakes tests produce more motivation than others. So what evidence is there for these assumptions?

2.4.1 The motivational impact of formulaic, narrow assessment

QCA's guidance on introducing stretch and challenge into examinations reveals the underpinning assumptions. Four of the five strategies suggested by QCA related to making the examinations less formulaic and broadening the skills being assessed: using a variety of stem types, avoiding atomistic assessment, using a range of question types and using synoptic assessment.

Examination predictability could take different forms. Predictability might be caused by the topics covered by a paper, the phrasing of questions, the structure and rubrics of the paper, question types

or the tasks candidates are required to complete (Ofqual, 2008). An element of predictability across these test facets is desirable. Question papers that change dramatically between examination series would have a negative impact on grade reliability, with test outcomes reflecting candidates' abilities to respond to unexpected questions as well as their subject knowledge – a form of construct-irrelevant variance (Crisp *et al.*, 2008; Ofqual, 2008; Wiliam, 2007). The line between desirable and undesirable predictability is hazy. It is therefore unsurprising that at present examiners receive little or no structured guidance on how to produce papers. As Murphy *et al.* (2007) argue,

'Examiners have to be extremely skilful if they are to succeed in producing papers that are sufficiently fresh to be interesting and challenging and sufficiently familiar to avoid provoking complaints'.
Murphy *et al.*, 2007, p23

The effects of 'over-predictability' are quite tangible however. If teachers were able to confidently predict the content of an examination, they might give students only restricted access to the specification content, choosing instead to prioritise 'test-worthy' content (Ofqual, 2008; McNeil and Valenzuela, 2001). Where candidates have been primed to respond to predicted questions they may be unable to unpick their carefully rehearsed responses and apply their learning to an unanticipated question, or question that is phrased differently (Ofqual, 2008). Indeed, candidates' expectations may be so strong that in the stressful testing environment they may not even recognise that a question is different from what they were expecting (Crisp *et al.*, 2008; Pollitt and Ahmed, 1999). Candidates might forfeit marks by not adapting their response to the question.

No studies have been identified that link student motivation with the above design features directly, but research on the National Curriculum Tests is, arguably, broadly applicable. For the National Curriculum Tests, there has been criticism of a formulaic, teaching-to-the-test approach (eg Wiliam, 2001) and there is evidence that students' views about the tests became more negative with age (Pollard *et al.* 2000). Hodgson and Spours (2003, p93) and Priestley (2003) also found that teachers reported teaching-to-the-test when *Curriculum 2000* A-levels were first introduced. Although there was no direct, empirical evidence relating formulaic and narrow assessment to reduced motivation, there is an appealing logic to the approach taken by QCA: a broad assessment should lead to wider learning and an element of unpredictability is intended to produce higher order thinking skills.

Broad assessments should have a positive washback, as they should encourage broader studying of the syllabus. Further, they should produce more valid assessments of students' knowledge and understanding. However, practicalities of time limitations in examinations could mean that the syllabus can only be sampled in a shallow manner. Thus, there is a tricky balance for examiners to achieve. An alternative approach would be to sample a topic in a deep manner, but make the examinations unpredictable in terms of the specific topics that would be covered in any one year. The art of question-spotting – predicting which topics will come up in the examination that year – then comes into play, but studying too little of the syllabus is risky for students under these conditions. Nonetheless, for modular A-levels, there are so many examinations that it would be natural for a pattern to emerge and for question-spotting to become an empirical matter. For students who sat the summer 2008 examinations, there had been 15 examples of individual AS examination papers since the first AS examinations in January 2001 and 13 examples of the A2 papers to learn from. Question paper design is not within examiners' gift entirely, as they must meet the Awarding Body's criteria and those specified by QCA. Oftentimes, examiners are highly constrained in terms of the structure of the question paper, style of questions and topic matter that they can assess.

Some evidence from a research report on examinations in Texas countered the idea that higher order thinking skills can be provoked by particular kinds of assessments. Gordon and Reese (1997) reported that students could be drilled to pass any kind of test, including those supposedly assessing higher order thinking skills. However, it was difficult to evaluate the quality of evidence presented by Gordon and Reese, because there was little data presented to justify their descriptions of their experiences of the impact of the curriculum and examination they were critiquing.

Unpredictable assessments could simply be less valid assessments. Examinations should be designed to produce an accurate picture of what students know and can do, but they might be caught out by a surprising question and prevented from displaying the knowledge and skills that they have. If we are seeking reliable tests of higher order thinking skills, then they already exist in the form of IQ tests (eg the Watson-Glaser Critical Thinking Appraisal; see Kline, 2000, p470). Test items for these assessments are kept secure, which means that people cannot necessarily prepare well for them. Thus, they meet the criteria of unpredictability and testing higher order thinking skills, which could lead us to think that we ought to be designing A-levels like this. But the conundrum is whether we should be trying to test intelligence through A-levels, or if we want them to be examinations that people can practise for.

2.4.2 The motivational impact of challenge and extended writing

Challenge and extended writing are intended to motivate broad and deep learning in the stretch and challenge policy. Sundra and Kitsanastas (2004) cited two papers, presumably conducted in the US², which showed that students were more motivated by multiple-choice tests than essay examinations (Jakwerth *et al.*, 1999; Zeidner, 1993). It is tempting to conclude that this preference for multiple-choice is a feature of the US context, but the causal explanation in the literature was that students were de-motivated by demanding examinations. From a study involving 301 high schools students in New Jersey taking a mathematics examination, Wolf *et al.* (1995) concluded that the 'mental taxation' (cognitive demand) of a question affected students' test-taking strategies. Two groups of students were compared: those for whom the examination results had consequences and those for whom it did not. Mental taxation was judged on a rating scale by a panel of experts. Ratings of mental taxation were not significantly correlated with empirically-derived measures of question difficulty and were therefore, arguably, conceptually distinct. Differences in the groups' performances on the examination questions were influenced by the degree to which the questions were cognitively demanding. Thus, the authors concluded that students' motivation was influenced by the amount of effort that they would have to make to get a question correct. Pintrich's expectancy-value theory of motivation has three components: 1) expectancy – likelihood of success, 2) value – importance of success and 3) affect – emotional reactions to the task (Pintrich and DeGroot, 1990). Wolf *et al.* (1995) added the impact of effort required to the value component of motivation in Pintrich's model. There is evidence that highly able students are motivated by challenge (eg Philips and Lindsay, 2006), but A-levels must be designed for a range of student abilities. Further, anxiety (Pintrich and DeGroot, 1990) and fatigue might affect even highly able students in examinations, such that it might not be the right environment for positive engagement with challenging material.

² These reports are not readily available.

To summarise, research findings were not encouraging in terms of the motivational effects of extended writing and demanding questions for students who are not high fliers. We might conclude from this that the stretch and challenge policy is pointless and not likely to work. But now that we have a generation of teachers who adopt teaching-to-the-test pedagogy (Gilbert, 2008), how would the Government influence teaching style without changing the tests? How could it be guaranteed that students would be stretched as part of the experience of doing A-levels, without designing that stretch into the assessments themselves? Indeed, Oates (1997, p146) promoted the idea that assessment modes should be selected on the basis that they foster desirable kinds of motivation, learning and achievement. There is a clear logic underlying the approach that has been taken, but alternative assessment possibilities exist. For example, Extended Project assessments have also been introduced, which allow students to pursue a topic in depth (for research on the impact of the extended project assessments, see Daly and Pinot de Moira, 2008). Let us turn next to what we know from the literature about assessing gifted students.

2.5 Assessing gifted and talented students

Provision for gifted and talented learners in England, Wales and Northern Ireland has recently extended to guidance for whole school and classroom initiatives on how to identify and support their learning and development (DCSF, 2007, 2008a and 2008b, CCEA, 2006 and 2007, ACCAC 2003). In addition, quality standards to aid the evaluation of interventions in the classroom and school-wide have been developed, (ACCAC, 2006, and Mouchel Parkman 2007a and 2007b). These initiatives are couched in terms of challenging, stretching and exciting learners so their potential is fulfilled, for example:

‘The main focus in secondary settings should be to create the right opportunities, with support and encouragement, to help the student to develop a desire to learn and sustain the personal drive that is required to fulfil one’s potential. This will be achieved by presenting students with work that challenges, stretches and excites them on a daily basis, in an environment that celebrates excellence.’

DCSF, 2007

Schools are thus receiving guidance on how to maximise the potential of learners in the gifted and talented subset, which will include high achieving A-level candidates, and are focussing attention on classroom strategies that overlap with the introduction of stretch and challenge elements at A-level. What can we learn from these initiatives that might help to make the most of the stretch and challenge elements at A-level? Specifically, what works when teaching and assessing gifted and talented learners?

It is useful to define what is meant by gifted and talented learners so their relevance to high achieving A-level candidates is transparent. A broad definition is “children or young people with one or more abilities developed to a level significantly ahead of their year group (or with the potential to develop those abilities)” (DCSF, 2008b). The definition is further broken down into gifted learners being those with high ability in academic subjects, such as Maths and English, and talented learners being those with high ability in creative and performing arts, sports, music or design. Given that stretch and challenge initiatives are intended for the highest achievers at A-level, the overlap between them and gifted and talented learners is obvious.

Evidence to support the use of some classroom teaching strategies was provided by a systematic review of interventions aimed at improving the educational achievement of gifted and talented learners undertaken by the Evidence for Policy and Practice Information and Co-ordinating Centre

(Bailey *et al*, 2008). Their review highlighted 15 studies in response to the review question: 'Which types of classroom-based interventions improved educational achievement of pupils identified as gifted and talented?' and were weighted according to their methodological robustness. Interestingly, none of the 15 studies were conducted in the UK and none investigate A-level performance. The reviewers also made the point that evidence was produced by teachers on 'what works' in many papers, but these did not necessarily meet the criteria of methodologically sound research, and would not be selected for such a review.

Assessment practices common to many gifted and talented programmes is the production of some form of creative product or a portfolio of work (Renzulli and Callahan, 2007; Sousa, 2003). Renzulli and Callahan state the case for the assessment of achievement from differentiated learning programmes, which advocate higher order thinking skills, creativity, investigative techniques and real-world contexts, to differ from paper-and-pencil tests and instead result in the production of a product. The exact nature of the product depended upon the programme and the expected learning outcomes, and they additionally provide guidance on their Student Product Assessment Form to aid teachers in the assessment of their pupils. One such product was a portfolio of work gathered by the student under the guidance of the teacher as evidence of learning (for example, Sousa, 2003; Johnsen, 2007). These methods have advantages in a differentiated curriculum, but teachers were less likely to opt for these authentic types of assessment when faced with high-stakes assessments and they were more likely to teach what might come up on the examination rather than assess a student's in-depth understanding (Hill, 2004, cited in VanTassel-Baska, 2007). Also, the use of a paper-and-pencil test offers the opportunity to cover all aspects of the courses content, whereas a product, such as a portfolio, is time consuming and does not have the same curriculum coverage (Damiani, 2004, cited in VanTassel-Baska, 2007). Whilst creative products offer advantages to assessing academic achievements of gifted and talented learners, they are not suitable for high stakes testing of school-based curricula because of the demands of timeliness, ease of management and reliability placed upon high-stakes tests. These issues mean that the Extended Project could not be the sole answer to the Government's quest for stretch and challenge of A-level students.

An answer to the dilemma of assessing the highly able within a curriculum-based assessment framework has been suggested by several gifted and talented specialists, such as VanTassel-Baska and Little (1991) and Renzulli and Callahan (2007), with the use of content-based assessments from a level higher than that of the student's age. Little and Ellis (1991) discussed the need to align programmes for the gifted and talented to the content standards set out within two particular secondary programmes: Advanced Placement (AP) in the USA and International Baccalaureate Diploma Programme (IB). AP offers 37 secondary courses and examinations in 22 subjects that students may take to earn credit or advanced placement at college (College Entrance Examination Board, 2009), and the IB offers a secondary curriculum in six academic areas, with the inclusion of an extended essay intended to encourage independent study by investigating a research question, collecting data and developing an argument (IBO, 2009). Little and Ellis (1991) argued that with training, teachers using these programmes could provide a learning framework that was rooted in a curriculum, offering challenging learning opportunities and providing assessment of content covered, and in the case of the IB, offered a creative product for students to produce. However, programmes for gifted and talented learners over the age of 16 would need to aim higher, such as at first-year degree level. This is clearly at odds with a common assessment for the entire A-level cohort.

Although there was research literature on the assessment of gifted and talented students, it focussed mainly upon identification of gifted and talented students, evaluation of special programmes or use of advanced curriculum materials beyond the normal syllabus. None of these research areas

generalised well to A-level examinations. Next, we look at what is known in the literature about students' views about A-levels.

2.6 Students' and teachers' views about A-levels

In a large-scale evaluation of the introduction of *Curriculum 2000* A-levels, Hodgson and Spours (2003, p89) reported that students initially felt over-assessed in the new regime, but began to see some of the benefits of modular assessment as they progressed through the course. They liked the fact that they could track their progress, as well as the possibility of re-sitting to improve their grades. In addition, the introduction of the AS qualification was welcomed because students could drop a subject after one year or leave school altogether at age 17 and still have something to show for their work. Students reported that they had to work very hard and there was not enough time to get everything done before the next set of examinations dawned. Interestingly, when this first cohort of *Curriculum 2000* students were asked what was needed to get through the A-levels, they said "motivation, perseverance and hard work rather than brilliance" (p92). Interviews with tutors indicated that the most able students felt stressed by their studies because they could not cover everything that they would have liked to study. However, by the second year of studies, these issues had reduced, at least for students who did not have too many re-sits.

Murphy and Warmington (2002) used a questionnaire to explore students' perceptions of the *Curriculum 2000* A-level examinations in 2002 and opinions of the media coverage of A-levels. Respondents were attendees at a University open day and had recently completed their AS examinations. When students were asked about their A-level experiences, a significant feature of responses related to exam pressure; respondents reported feeling that they had to learn to just pass the exams, and there was no chance to discuss things fully. Students also felt that they spent too much time being examined, and that the examinations themselves were more important than their experiences of the courses. Media reports suggesting that standards were falling made students feel demoralised. Students reported studying hard during their A-levels, but their efforts were often undermined by the negative media surrounding the publication of examination results. Another questionnaire study (Baird, Ebner & Pinot de Moira, 2003) explored students' choice of study following curriculum reforms in 2000, as well as their perceptions of the media reporting of examination results. Whilst media reports commonly suggest that A-levels are becoming too easy, only 1.7 *per cent* of those surveyed believed their AS examinations were too easy, and 48.3 *per cent* felt that they were too difficult. Student views clearly contrasted with media reports: students believed that they put considerable effort into their examinations and often felt disheartened by the negative press.

Ipsos MORI have conducted surveys of students', teachers' and other stakeholders' views of the examination system on behalf of QCA annually since 2003. In the most recent report (Ipsos MORI, 2009), the vast majority of teachers (95%, $n=332$) and students (90%, $n=88$) considered A-levels to be important qualifications. Since 2003, teachers' agreement with the statement 'Most students taking A-levels work hard' had not changed, with 71% ($n=332$) agreeing in 2008. Teachers with more than 25 years' experience were more likely to consider that students were working hard than those with five year's or less experience (p23).

The above research set the scene for the current study, which was focussed upon students' views on the examinations themselves, as well as their views on A-levels generally.

2.7 The current study

This research set out to investigate students' views of the new, specimen A-level question papers, which had been designed to incorporate stretch and challenge. One problematical aspect of any trialling research with examination papers is the sample of students to be included. The purpose of the research is to find out how the design of the assessment will affect students, but there are naturally other factors that impact upon students, such as their readiness for the examination. Ideally, students would have given their views on the trial papers at the same time of year and with the same level of preparation as students who will sit the live examinations. For the stretch and challenge A2 question papers, this would not be feasible until 2010, when the first A2 examination papers for the new A-levels will be sat by students who have completed two year's study of the new curriculum. Thus, if anything is to be known in advance of the first live examinations, some compromises had to be made regarding the sample of students who were involved in the work.

Another problem for many examination paper trials is recruiting sufficient participants. With the introduction of modular examinations, the opportunity to practise examination technique is no longer an incentive and awarding bodies openly publish examples of assessments, so there is not much additional information to be gained from participation in a trial even when examinations are new. OCR's (2007) trial of stretch and challenge questions in psychology, chemistry and English literature suffered from this problem, as they only recruited 10 *per cent* of the participants that they sought.

To address the above sampling problems, participants who had completed A-levels in 2008 were recruited to participate in the study. These participants had experience of the full curriculum and their A-level results and experiences were fresh. A disadvantage of this sample was that they did not have experience of the new curriculum, but as already indicated, it would not have been feasible to recruit a sample that did have that experience without waiting until 2010. Psychology and biology students who sat A-level examinations in 2008 were recruited by e-mail from the universities of Bristol and Manchester. (Appendix A)

3 Methodology

3.1 Sample

A purposive sampling approach was adopted, in which grade A students at the universities of Bristol and Manchester were targeted, with a focus upon two subject areas: biology and psychology. These subjects were selected to ensure that there was likely to be a wide pool of students who may be attracted to the study. Following the initial advertisement, 45 University of Manchester students and 28 University of Bristol students volunteered.

From the volunteers, students were prioritised for recruitment on the basis of their examination results and availability for interview. The aim was to include 10 students from both subjects in each of Bristol and Manchester, giving a total of 40 interviews. Students were over-recruited to allow for attrition. However, only 39 interviews were achieved in practice, due to students dropping out (despite proactive strategies – see Procedure section below). All males were prioritised because there were so few male volunteers, but the final number of male participants was only five (34 female). Biology students proved more difficult to recruit than psychology students (Table 1), so four students who were studying degrees in psychology were interviewed about their experiences of sitting a biology

A-level the previous summer. Independent Schools were over-represented in the sample, which is to be expected of a sample of grade A students (Table 2), as A-level examination statistics demonstrate (http://www.jcq.org.uk/national_results/alevels/). Efforts were made to ensure that the sample reflected experiences of more than one awarding body for each subject. Students had experience of four awarding bodies across their A-level subjects (AQA, Edexcel, OCR and WJEC), although in one third of cases, students were unable to identify the awarding body that certificated their A-level qualification. In the subjects being studied (psychology and biology), no student had taken their A-level with WJEC and most students had taken the A-level with AQA. A geographical spread of students' schools was achieved. Most of the participants had been awarded a grade A in the subject under study and seven students had been awarded a grade B.

Table 1 Subject under investigation and university attended

A-level subject	University of Bristol	University of Manchester	Totals
Biology	8	10	18
Psychology	5	16	21
Totals	13	26	39

Table 2 Schools and colleges of origin for the students in the sample

	Manchester (%)	Bristol (%)	All (%)	National %
Secondary Comprehensive	11 (42.3)	4 (30.8)	15 (38.5)	33.05
Secondary Selective	2 (7.7)	1 (7.7)	3 (7.7)	2.58
Secondary Modern	0	0	0	3.20
Independent School	3 (11.5)	5 (38.5)	8 (20.5)	12.31
FE establishment	0	1 (7.7)	1 (2.6)	7.46
Sixth Form College	10 (38.5)	1 (7.7)	11 (28.2)	1.64
Tertiary College	0	0	0	0.81
Other (inc private candd)	0	1 (7.7)	1 (2.6)	38.96
	26	13	39	

3.2 Materials

Semi-structured interview schedules were constructed to reflect the three main research areas of the project; experiences of GCE, perceptions of examination standards, and perceptions of the specimen stretch and challenge papers (Appendix B). However, in order to understand the diversity of students' experiences, interviewers were encouraged to digress from the schedule and follow interesting issues raised by participants. Initially, the interview focussed upon students' recollections of sitting the particular examination paper under research. Then the interview moved on to contrasting the questions on the examination that the student had sat in summer 2008 with the new style of question paper incorporating stretch and challenge. Students were presented with summer 2008 question papers for the A-level that they had sat that previous summer (from the relevant awarding body).

They were also shown specimen examination papers from the same awarding body for the new A-levels which were designed to incorporate stretch and challenge. Finally, students were asked about their views of A-level examination standards.

The two subjects included in the study were selected because they had high numbers of students and therefore recruiting participants would be feasible. Additionally, the psychology syllabus had undergone a structural change, as it would be assessed by four units in the new examination rather than six. Biology, on the other hand, remained unchanged, with six assessments.

3.3 Procedure

Advertisements were sent by e-mail to students, indicating the nature of the study and sample being sought, as well as announcing that a fee of £40 in book tokens would be paid for an hour's interview. This was deemed an appropriate incentive as students were in their first term of university and would be purchasing textbooks. Volunteers were asked to give contact information and some details of their school and A-level results via a web form constructed on SurveyMonkey (<http://www.surveymonkey.com>). This proved to be a very efficient way of recruiting and selecting participants. Once participants had been selected, they were contacted by telephone to confirm the interview arrangements. Participants were also contacted on the morning of the interviews for confirmation of attendance. Despite these procedures, initial drop-out rates were high and some interviews were re-scheduled and replacements were sought.

Five interviewers were involved in collecting data for this research. Interviews were conducted on a one-to-one basis, with the interviews being audio-recorded and professionally transcribed by a transcription company. Field notes were taken by the interviewers and these were provided to the researcher who conducted the analysis. Each interview lasted between half an hour and an hour. Students signed consent forms and were given information about the study to take away. Ethical approval for the research was provided by the University of Bristol and the research conformed with AQA's and QCA's Research Codes of Practice. Consent forms were signed and kept on record and participants were given an information sheet to take away, which included instructions on how to make a complaint about the study.

3.4 Analysis

The qualitative data were transcribed verbatim and analysed using dominant theme analysis; one of the more straightforward qualitative analysis techniques for applied social research (Maxwell, 2005). This involved several readings of the transcripts, while searching for and checking the validity of dominant themes. On later readings the participants' comments were coded using nVivo software. Some codes were produced from the memos provided by the interviewers and others were generated in response to the data. The dominance of a theme was defined by its extensiveness – the degree to which similar comments were made by several people – and the strength of feeling for a particular topic. Participants' comments are used below as evidence of the dominance of each theme.

Coding was conducted by one of the interviewers and then a validation process was undertaken. With such a large corpus of qualitative data it was necessary to explore the dependability of the analysis without repeating all of the coding work. A day-long meeting was held in which three of the research team members who had not conducted the initial analysis were given clean and un-coded

copies of five transcripts. One transcript was selected to represent each of the five interviewees. Overall, the five sample transcripts covered the themes identified in the initial analysis.

The validating researchers read each transcript, noting themes as they emerged and mapped the evidence in the transcripts to the themes. Following this, the original analyst's themes were revealed to the group and a group discussion ensued. The aim of the discussion was to negotiate the dominance and scope of each theme, grounding this in the evidence within the transcripts. As a result of the discussion, some of the themes were re-organised. Following this meeting, the findings were sent to another of the interviewees for validation. This method therefore addressed dependability and subjectivity issues, as the final check on the coding process was conducted by a researcher who had not been involved in the discussion of emergent themes and had an independent view of the data.

An initial analysis investigated whether there were differences in the findings between Bristol and Manchester, and between subjects but no general differences were found. Thus, the data were collapsed into a single analysis.

4 Findings

The first sub-section below reports students' responses to the stretch and challenge specimen papers. Their responses covered the layouts of the new papers, question type and content coverage, and the students' perceptions of demand. Three other main areas were covered by the interviews, one of which had a substantive sub-theme:

- 1) Examination nous;
 - a) Aspects of difficulty;
- 2) Motivation and stress among high achievers;
- 3) Engagement and challenge in examinations.

Each of the above three themes is outlined in detail following the description of students' reactions to the new question papers.

4.1 Students' reactions to the stretch and challenge specimen papers

During each interview students were given the opportunity to read and compare their actual paper from 2008 and the relevant version of the stretch and challenge specimen papers. This section details the students' responses to the specimen papers in terms of their layout, question type and content coverage, and perceptions of demand. A complex pattern emerged because in each case, students were divided in their reactions to the new question papers. This is perhaps best illustrated by their reactions to the last section of questions in the interviews, when students were asked to rate the perceived demand of their actual and specimen papers. Table 1 shows that there was an even split in responses, with some students perceiving the new style of question papers to be more demanding and some less, and approximately the same number perceiving the demand to be equal. The inconclusive distribution of responses was consistent with the range of comments made by the participants. This could be in keeping with the overall policy aims as stretch and challenge was not intended to be more difficult for students and, for individuals, demand and difficulty are one and the same. However, it could reflect variability in the success of introducing more difficult questions in the question papers presented to students in this study.

Table 3 Ratings of specimen papers compared to summer 2008 A-level question papers

	More demand	Same demand	Less demand	Total
Biology	6	4	4	14
Psychology	6	6	7	19
Total	12	10	11	33

4.1.1 Layout

Some students liked the layout of their specimen paper and thought they looked straight-forward, uncluttered and well-structured. This made it easier to spot the topic areas they intended to attempt. However, some thought that the papers looked more cluttered and wordy. The following students appeared to interpret 'wordiness' with a greater demand for comprehension skills.

'[If I'd had the new question paper] I'd have been angry because I'd revised for the psychology exam and I knew a lot about psychology but I wouldn't be able to do this because I'm not good at English. I think I'd be quite angry because someone who was good at English and maybe they didn't know as much, but because they could understand the question they could get an A and I'd get a C...I really hate these [new] papers.'
Psychology student

'The wording doesn't seem different, just more. It looks heck of a longer exam. It doesn't look like you could flick through it and go 'there's the questions I can answer, get them out the way'. This one you've got to read everything and I tend to avoid reading things and just get that information down.'
Biology student

4.1.2 Question type and content coverage

Most students commented that the specimen papers contained more questions and that these were of the short answer type - this observation applied to both subjects but especially psychology. Individually the questions tended to be worth fewer marks than those in the legacy question papers, particularly in psychology where the choice previously tended to be between essay questions worthy of 25 marks. With more questions, the new question papers were deemed to have a broader content coverage. The (perceived) broader content coverage had implications for learning and revision and some students felt that they would need to know and revise more of the specification content. This clearly has links to the stretch and challenge policy, as the following quotations demonstrate that students considered that they would have to study more broadly. Note that there was no indication that any more depth of analysis would be required by the new-style question papers however.

'With the new paper you'd have to learn a lot of things rather than a set of basics.'
Psychology student

'You'd be more challenged in that it looks like you need to know a broader content. Like, you wouldn't be able to like just revise one [topic] and then hope that it came up.'

Psychology student

'I think it would have made the revision more challenging because it's quite specific. You'd need to know quite a lot of detail.'

Psychology student

Some students liked the fact that topic areas were broken down into shorter response questions in the new-style examinations. This meant that they could pick up marks quite quickly and easily, and would not have to devise a structure to their response (as it was already broken down for them), as illustrated by the following quotation.

'If you look at a question and it's 30 marks and you don't know anything about it, you panic and it's a lot harder. Whereas, say, like a question that's then broken down into four parts, although it could still be asking the same things, the guidance just makes it a bit easier'

Psychology student

Conversely, some were concerned that the shorter response questions required a 'right' answer (unlike essay questions) which would increase the likelihood of getting some of them wrong. Some preferred essay questions as it gave them the opportunity to just write everything they knew.

'If you get 'describe' they just let you reel off a load of information. They're really good questions and I can't see any of them [in the new paper] (laughs)!'

Biology student

The fact that each topic area consisted of more questions complicated the question choice process. Students commented that they would have to weigh up how many questions they could answer in each topic area. This was exacerbated by different topic areas being assessed by different question types. A student might prefer the content of one topic area but might feel she could pick up more marks in another topic due to the question type. Choosing between questions was a real dilemma and source of uncertainty for many students, and making the decision could be stressful and time-consuming. One student commented that she would have to map out her answers to all questions to determine on which one she could score the most marks. Another suggested that,

'I would have thought, right, even if I knew more on Cognitive Behavioural Therapy I want to do Anxiety Disorders 'cause it's easy to get, like, smaller marks than it is to aim for a twenty-five [mark question].'

Psychology student

Some students commented that the earlier question papers had indicated where a synoptic response was required with an 's' and that they would be afraid to write a synoptic response if it was not made clear that it was required. A biology student commented as follows,

'I don't know if I would personally feel confident putting in synoptic [writing] into a paper that didn't sort of ask for it because I would maybe feel I wasn't answering the question in the way they wanted.'

Biology student

Requirements for synoptic writing clearly need to be signalled to teachers and students if stretch and challenge is to be a success. Many students remarked that the specimen papers contained a larger number of applied questions. This was identified as a significant change, and many had not seen such questions before or would know how to tackle them:

'For instance, 'Create your own brief moral comparison story', I don't think that we had to sort of make up our own stuff. We didn't have any questions like that, but I think it's good.'

Psychology student

'They've given you a new experiment to look at. I've never seen that before. We would never have prepared for anything like that.'

Psychology student

'The first question, it's like beyond challenge. It's just so abstract. Like, we would never have covered this before. I can't interpret the diagram.'

Psychology student

'I don't know how you can teach applied. I don't know how you can learn how you apply – you either can or you can't.'

Biology student

'[This question] is something I haven't really seen before. There's like a synoptic part to it and it's not something you'd be prepared for or have a model answer for. It's 'Explain why policymakers would be more inclined to accept the evidence of these published studies rather than their opinion'. It's talking about the credibility of clinical research which is something you'd be aware of but you wouldn't have a model answer for. I'd feel confident doing the question but I'd say it's slightly more challenging.'

Psychology student

Many students suggested that the applied questions would be difficult to prepare for (as they would not know the exact details of the case study etc) and that this could increase stress and uncertainty. However, some thought the applied questions were a better test of their knowledge, and could encourage a more creative approach in the examination, as the following quotes suggest:

'I know some people that like, well, they didn't like [studying for their exams] very much, but they got by with just learning everything and not necessarily understanding it. This [new paper] sort of tests your understanding as well as learning.'

Biology student

'I think you'd have to work harder. It's kind of thinking for yourself, instead of just memorising notes. So, yeah, pulling all your knowledge together, I'd find that difficult.'

Psychology student

'This [new paper] is a lot more practical than the sort of papers I've sat, maybe a bit more scientific. There's a bit more terminology in there. I think it's a step forward. Maybe harder to revise for. Probably more enjoyable in the grand scheme of things.'

Biology student

'It's more challenging because you can't learn the answers.'

Biology student

'I think this [new paper] tests a different skill. It doesn't test what you know about biology, it tests how you use your biology.'

Biology student

'I've got a practical today where I'll probably get, you know, some stuff to look at and write up, make my own conclusions from, which is what this paper is trying to get at.'

Biology student

Overall, then, students did not perceive the new-style question paper to be more demanding generally, but the applied questions were perceived as being more challenging and as making

students think deeply in the examination. Further, the structuring of questions induced a perception that there would be broader content coverage. Perhaps jointly, these features of the examination could produce more stretch and challenge.

Wider issues were discussed with students in the interviews and those findings were just as interesting as those directly related to stretch and challenge. It is the themes of those findings that we turn to next.

4.2 Examination preparation

This was a powerful theme emerging from the interviews, involving issues that were clearly of great significance to the students. Students typically recounted emotively the extent to which preparation for their examinations had dominated their lives over the short and longer term. In particular, students spoke about how they were trained to perform in examinations by their teachers, whilst noting some of the limitations of highly strategic approaches to learning and revision; how past papers and mark schemes formed the foundations of this training; and how students' examination training experiences were underpinned by a perception that question content and demand are predictable over time. The majority of interviewees recalled receiving intensive training from their teachers in how to prepare for an examination and respond to questions. Typical examination techniques that teachers trained students to use are given in Appendix E.

Students learned and practiced these skills mostly by using past papers and referring to mark schemes. These were used to deconstruct questions, evaluate the merits of model answers, practice responses and receive teacher feedback, and understand what examiners were looking for (see below for further discussion). The students had practiced with so many past papers (in some cases all of the past papers that were available, either in class or alone) that they were often unable to identify their own summer 2008 examination paper during the interviews because all papers looked familiar. Even the new specimen question papers had familiar-looking questions, as the following two quotes illustrate.

'We've done this. These exact questions, the whole paper. That's strange, I can remember this paper.'

Biology student

'I've done this paper. This so rings a bell.'

Biology student

In contrast, some of the students believed the new papers to be very distinguishable from the legacy papers. As the discussion of the students' responses to the stretch and challenge specimen papers demonstrates, many students had not previously encountered questions requiring the application of knowledge, and were unfamiliar with the style of questioning used in the specimen papers. It is entirely possible, given the time-scales for constructing the question papers (Baird and Lee-Kelley, 2009), that examiners did base their specimen question papers on previously published questions.

Past papers were a very important means of demonstrating that examination papers were consistent in terms of layout, rubrics, the distribution of marks and most importantly, topics and questions. Knowing this was essential for the students' levels of confidence and their expectations, as the following quotation demonstrated,

'We had so much exam practice that no question could really surprise us.'

Psychology student

Students had a clear expectation that examination questions would be like those that had gone before – something that has not been true historically of examination question paper design, but has been the case for at least the history of Curriculum 2000 A-level examinations. Students spoke about some of the other efforts made by their teachers to help them predict questions and prepare for their examinations. These are listed in Appendix E.

Students found it reassuring to feel that they could predict the content and coverage of each question paper, and found it valuable for structuring their learning and revision. However, the goal of predicting questions was also constraining, as it closed avenues of learning, encouraging formulaic responses and discouraged students from applying their learning in different contexts. The experience of the following student was not uncommon and illustrates the risks inherent to the process of predicting questions,

'The teacher had told us to predict [the questions] 'cause he said there's always a routine and you can always guess, but he didn't guess it right that time. I still got an A but I nearly started crying when I saw the questions'.

Psychology student

Examination training started early and pervaded the two years of A-level study by underpinning the subject-based learning. For example, some students went into their exam knowing that they could only answer questions numbered two and five, and that they had not been taught the content required to answer questions numbered one, three and four. Having a clear idea of which questions they would answer saved time during the exam and reduced anxiety associated with choosing between questions. The interviewees were mostly very concerned that they demonstrated their knowledge and learning in responding to the questions. A few were disappointed that they could not show more of their learning, or thought that the examination was an anti-climax, testing only a small proportion of their knowledge. Given that the sample consisted only of high-achievers, the views of the following two students may not be held widely among the student population:

'I was quite gutted when the time was up because I had so much more to write.'

Psychology student

'You worry about all the different types of questions they could ask you but sometimes it's an anti-climax. You've prepared reams and reams of information and they ask the question in its simplest form.'

Psychology student

Many commented that good performance depended, at least in part, upon understanding the mark scheme, and knowing what knowledge was valued by examiners. The following quotes illustrated the extent to which students relied upon mark schemes to guide their learning, revision and performance in the examination:

'That's the worst, when you just have to say a phrase [to get the mark], so then you just end up learning the phrase and putting the phrase in your answer rather than learning the answer itself.'

Biology student

'You looked at the mark schemes and you had to figure out how [the examiners] would be thinking and what the logical answers were.'

Biology student

'I would tell anyone just to focus on the mark schemes and past papers and go through them more than the content because they teach you a lot more.' Psychology student

'You need to understand how you need to answer the question and what the examiners are looking for. That's really important 'cause if you just learn the material and not understand how you need to answer the questions, you won't do so well.'

Psychology student

There appeared to be three drawbacks to the examination training approaches that were based on predicting questions and the use of past papers and mark schemes. First, they encouraged formulaic responses and discouraged individualism, as the following quotation demonstrates:

"Everything was very formulaic which was quite annoying because we just spent all year sort of learning how to answer the questions more than anything else. No one really wanted to take any risks in their exam results."

Psychology student

The approaches to revision and the topics being revised had proved successful in the past and students and teachers appeared reluctant to change their preparation strategies. Even students who would have liked to focus on another topic were afraid to take risks or deviate from what had been successful. Second, reliance on past questions, papers and mark schemes superseded the need to understand the content in some extreme cases. Students sometimes focused upon learning key words and phrases:

'I literally learnt the mark scheme. I was like, well there's no point in trying to go into the details of why this [biological process] works. I knew exactly what wording they wanted.'

Biology student

The same interviewee also suggested:

'Like, don't just do a past paper once. Do it again and again and again until you know every answer, like until you're near 100% 'cause your likelihood of the questions coming up is quite high. The questions are really similar so you're bound to hit a few that you know the mark scheme for.'

Biology student

Superficial, rote-learning of marking schemes and model answers was a worrying feature of some students' experiences of studying for A-levels. Third, many students struggled to adapt their planned responses in the face of an unanticipated case study, scenario, quote, diagram, command word or phrasing, as the following quotations illustrated,

'My teacher told us they ask this question in exactly the same way every year so she gave us like an essay plan for that question. I really learned it, in memory. I was so proud of myself 'cause all my friends were like 'I hope it doesn't come up' and I thought 'I do!' It was on the paper but they'd worded it differently, like the AO1 and AO2 were different, and I was like 'oh no, what the hell?' because I didn't know how to break that essay plan up into like a new one.'

Psychology student

'In our year [the writers of the paper] decided to whack in this extra space thing which completely threw everyone off. They'd never put in that extra space thing before and we got into the exam and everyone was like 'why is there extra space?'

Biology student

Given the high stakes nature of A-levels, it is appropriate that students received coaching in how to succeed. Interestingly though, many students recognised the limitations of these approaches and the fact that they diverted their attention away from learning the subject content. This fundamental issue will be returned to in the discussion section.

4.2.3 Aspects of difficulty

Many students perceived the issue of question difficulty to be intertwined with examination preparation. As well as more traditional conceptions of question difficulty, questions were considered more or less difficult based on the quality of the preparation. The following quotation was indicative of many comments on this issue:

'What makes an exam question easy or difficult? Whether I know the knowledge or not. If I've learnt it then it's not going to be hard.'
Biology student

Many students commented that the perceived difficulty of the first question had a strong influence on their confidence and performance throughout the rest of the exam. An easy first question had a settling effect. Conversely, a difficult first question made students feel panicked. This hindered their ability to read the other questions calmly and clearly, increasing the likelihood of making false starts and complicating the process of choosing between questions. Under these conditions the students also reported difficulties in recalling information and answering the questions to the best of their ability. Students did not always respond well when faced with something unanticipated as the following quotation demonstrated,

'The first question was on cells or something. It actually sticks in my head 'cause I couldn't do it. I couldn't understand the diagram at all. I was absolutely gutted when I turned the page and didn't understand what the hell was going on. I panicked and thought 'I can't do this now' and I walked out. Everything else is a haze.'
Biology student

Clear command words, such as 'describe' and 'describe and evaluate' or 'outline and evaluate' were preferred. Students were confident that they understood what these command words required of them, and could give examples of how they would structure such responses. Questions based on the command word 'discuss' were less favoured as it was suggested that this gave little indication about the appropriate structure for their responses, as a psychology student indicates,

'I need to know how the marks are distributed. Like, how many for 'outline', how many for evaluating. This gives me an idea of the number of points I need to put in the text of each section.'
Psychology student

Application of knowledge (with the use of a case study for example) was largely perceived as more difficult, both in terms of preparation and during the examination. Indeed, given the thorough approach to examination preparation that most students had encountered they appeared surprisingly unprepared for applied questions. The majority were not keen on applied questions (due to their higher demands) but a small number thought they constituted good preparation for higher education and were suitably challenging. The following quotations illustrated both attitudes.

'My revision technique really worked until I had to like apply my own knowledge, and then it didn't.'
Biology student

'[Applied questions] separate those who know from those who understand.'

Biology student

Almost all students remarked that while they had found their subject interesting, they did not find the examination paper interesting. This was mainly because they were not thinking of a paper's intrinsic interest while attempting to complete it, which does not bode well for attempts to engage students in examinations. Indeed, we could question whether it is an appropriate function of an examination to engage students. Given these experiences, the discussions around what motivated students' learning were interesting. We turn to that theme next.

4.3 Motivation and stress among high achievers

Interviewees were motivated to succeed in their A-levels primarily to meet the entry conditions for their chosen universities. A small number of interviewees were motivated to achieve three A grades in order to apply for scholarships for high achievers. It is tempting to interpret the instrumental nature of students' learning experiences as indicating an easy life, but we must disabuse readers of this notion, as students reported hard work, stress and anxiety in relation to the examination results, as these students noted,

'When you look back it seems like a lot of worry for nothing when you get your grades, but, when you don't, I can't even imagine what it would feel like because I know how stressed out I was beforehand. I was really anxious, and all I wanted was to [study at the University of Manchester]. If I hadn't have got in [to the University of Manchester] I would have been so upset.'

Biology student

'I think everyone just wanted to get grades to meet their offers really. I don't think many people were particularly interested in sort of achieving for achievements' sake by that point, because we were all a bit tired.'

Psychology student

'My Dad rang [me] and said 'You've got As' and I was just like, 'pew'. There was no, sort of like, 'YES, I've got As!'. I was just like, okay, sorted.'

Psychology student

The last quotation illustrated a phenomenon evident in this high-ability group: getting good results was about ticking a box. Expectations of good results were the norm and anything less would have been a surprise. Some students responded with horror to notion of the introduction of the A* grade – the notion that there might be even higher hurdles introduced and that a grade A would be devalued – for it was not that the examinations were effortless, simply that with intelligence and hard work, students could expect to get into the club with the best results. Creation of an even more elite club would have increased the pressure.

Interviewees were more likely to want to please or prove something to their teachers than their parents. Many students reported good relationships with their teachers, suggesting that their teachers had made special efforts for them or been particularly encouraging or supportive,

'The tutors want their pupils to do well. I know, I can honestly say, with my psychology teachers it wasn't like they were trying to get me to pass so I could be a statistic. They actually wanted me to do well.'

Psychology student

The majority of students reported that their examinations induced feelings of anxiety, stress, nervousness and fear. In the more extreme cases students said they felt petrified and traumatised.

At least one student had missed examinations through stress. Of course, these students were a biased sample, as they had experience of examination success and it could be that examination stress is worse among students with experiences of failure or partial success. The importance of the examination result, examinations scheduled closely, question choice, ambiguous questions, lengthy text, difficult first questions, uncertainty about how to structure an answer and the formality of the examination environment were all mentioned as features of the examination experience that were stressful.

4.4 Challenge in examinations

Every student interviewed felt strongly that exams were not getting easier over time. Indeed, as they were so familiar with past papers for the *Curriculum 2000* examinations in their subject they felt they could make good judgements about demands. Of course, this does not allow for the standard-setting process and is a perception of the question papers alone. We have included several student comments below, to allow their voices to respond to the allegations.

'I thought [my exams] were hard enough and we all struggled. No one just sailed through. So I think they are challenging and they're not getting easier. I think people are just getting better at teaching them and helping people to get the right answers.'

Psychology student

'I don't see them getting easier – I just see other factors making them slightly easier for certain people.'

Psychology student

'I just feel that [my exam papers] were appropriate for my level – the right amount of content, the right style of questions and coming out with a reasonable knowledge. I've gone into a biology degree feeling reasonably prepared.'

Biology student

'Maybe we've got better resources than they did, and maybe we've got access to the internet and loads of books and different methods like revision podcasts and whatever, that they didn't, but it doesn't make it any easier. If anything it's harder because there's all the new [content], like DNA. They didn't have to deal with any of that twenty years ago.'

Biology student

'I went to a good school and I consider myself fairly bright but I didn't find it easy. Everyone at my school worked really hard. I feel like when you get all the backlash it sort of detracts from the achievement because everyone does months of revision and they spend a lot of time working towards good grades.'

Psychology student

As examinations were not perceived to be easier, the students offered two explanations for rising pass rates. The first was that there was greater competition for university places which forced students to work harder to achieve the required grades – the qualifications arms race described by Wolf (2002). The second explanation was that students and teachers were more informed and experienced, especially as they were able to re-sit examinations. They suggested that teachers improve their teaching methods year on year as they became more familiar with the specification and have more experience of success. Students also noted that they had greater access to better information. Effectively, this second explanation related to learning about the assessment, rather than the subject.

Some students were aware of the impact that performance tables had upon outcomes, as the following two quotations illustrated:

'Say my college, last year 80% got a pass. The governing body or whatever will turn round and say 'right, this year we want 82% to pass'. So you get that 82% and the media look at it and go 'each year they're getting easier'. The better the college does, the higher the targets for next year.'

Psychology student

'There was at least three people in my chemistry class that would have failed A2. They got a pass at AS but if it looks like they're going to drop off the edge at A2, then it's like don't come back because we don't want, you know, we don't want the fail. We got a 100% pass rate.'

Biology student

Increasing the challenge of A-levels was considered unnecessary, unfair, and demoralising. Indeed, as the following quotations illustrated, many students were angered by the suggestion that examinations should or could be made more challenging:

'Because more people can understand the techniques needed to take an exam and more people are passing, everyone's complaining. But why are they complaining when everyone's getting good marks? It seems quite a silly thing to do. Why would you want to make it so that less people can understand and get a good mark?'

Biology student

'Everyone's putting pressure on you for your future and then people turn around and go 'what you're doing is easy, it's pointless' I don't think people should sit there and say that what I've done, what I've spent literally my whole school career doing, is easy.'

Biology student

'I don't think you should make papers harder. You're just going to put people off if they can't do any of it. That's just depressing.'

Biology student

Two students suggested that there is an important distinction between being challenged in the classroom and the examination:

'I think they should all be challenging because it's an A-level and you need to be challenged, and it's hard and you need to prepare yourself for university. But in the same sense, if you've done your revision, it should be challenging but answerable, so the challenge should all really be before the exam.'

Biology student

'You don't really want to be challenged [in an exam]. You just want to get down what you know. I think you should be challenged in class. In an exam your brain can be a bit everywhere.'

Psychology student

5 Discussion

Innovative teaching and learning methods were experienced in abundance by these students, yet their approaches to study were formulaic and grounded in rote learning. Past papers and mark schemes played a central role. This formulaic approach was facilitated and encouraged by teachers

and was entirely rational and even creative in some instances. Teachers and students were driven by grade goals and the strategic approach ensured some level of success. Deviating from the past papers and mark schemes would entail unnecessary risks. Even students who were frustrated by the constraints of this approach conceded that it had merit and was functional. We are creating a culture of professional learners.

In the group of students interviewed, there was a great deal of engagement with the subject matter. Students had worked very hard, but had not necessarily been stretched or challenged by the examinations themselves. They questioned whether examinations were the right place for that to occur, but also recognised that the examinations had interfered with deep learning. The psychology students in particular had learned how to deal with extended writing, which would traditionally be considered a higher order skill, but their approaches were formulaic. As such, this research adds to the literature by demonstrating that students can be drilled to produce essay-style answers, as well as multiple choice responses. Ironically, some of the students were being assessed by multiple-choice examinations in their first year at university and could not use their essay skills as much as they would have liked.

Students offered many interesting insights into what made an examination paper difficult. The high stakes nature of the testing environment and the need to meet university offers are significant sources of stress, but there are many other contributing factors. At the level of the individual question paper these include the difficulty of the first question, the chosen command word, the ambiguity of the language used, uncertainty about how to structure a response, use of excessive text, complicated or vague diagrams, and having to choose between questions. Question writers should be aware of these sources of stress, and consider whether each of these is a legitimate means of differentiating between candidates. That is not to say, however, that designing these features out of examinations is an easy process, as the research outlined earlier indicated.

Overall, students did not perceive a large shift in stretch or demand of the new A-level specimen question papers. However, many observed a shift towards a greater requirement for the application of knowledge. On the one hand, such questions were considered to be difficult to prepare for and could potentially increase stress and uncertainty. On the other, they were seen by some as fun, interesting, challenging and relevant. Secondly, they observed greater use of short answer questions. Some perceived these to be easier point-scoring questions, while others believe that they will demand a broader knowledge base. Many students suggested that these changes would challenge the usefulness and relevance of established means of learning and revising. As such, it is possible that some of the changes made to the question papers will result in the intended stretch and challenge, through washback effects upon teaching and learning. Equally, teaching and learning might adapt quickly to accommodate the style of questions and students might simply be drilled in different ways than before.

Earlier studies, conducted when *Curriculum 2000* was being introduced, had indicated that students were working very hard to achieve their examination results, to the exclusion of enrichment activities (Hodgson and Spours, 2003; Priestley, 2003). Students in the current study reported similar experiences, which demonstrated that the earlier results were not simply due to teachers over-egging the curriculum content of the AS in the first years of its introduction. Examinations were a dominant feature of these students' learning experiences and they appreciated that they detracted from learning the subject content. Sadler (2007) called upon examination boards to show leadership in this area, writing,

'I believe that a blinkered conceptualisation of curriculum, the strong trend towards fine-grained prescription, atomised assessment, the accumulation of little 'credits' like grains of sand, and intensive coaching towards short-term objectives, are a long call from the production of truly integrated knowledge and skill.' Sadler (2007, p392)

Underlying this area is a deep philosophical issue about the nature of knowledge and the purpose of education. Dweck (2006) wrote about two different approaches to learning – fixed and incremental. In a fixed mindset, the approach is taken that achievement is driven by ability, whereas in an incremental mindset, effort produces the outcomes. It would be perverse to argue that we would not wish students to work hard to gain their A-level results, or to withhold the cultural capital that they need to succeed by way of explaining what kinds of knowledge are credit-worthy. We would surely wish to encourage incremental mindsets. Yet the current approach to this has led to the situation so aptly described by Sadler. Will the adaptations to the A-level question papers be enough to engender deeper, more integrated approaches to learning?

6 Conclusion

There are some encouraging signs that the new-style, stretch and challenge examinations might have a washback that fosters more flexible thinking and broader study. This research was conducted in only two subjects, but did cover the three English Awarding Bodies' specimen question papers. However, some caution is necessary. Participants in this research had not studied the courses that will be assessed by the new A-levels and was focussed upon students' opinions, not their actual performances in examinations. Further, it is possible that it is not the nature of the questions themselves that have brought about the undesirable kinds of teaching and learning described here and in other studies, but some other aspect of the *Curriculum 2000* A-levels, such as their modular nature (Hayward and McNicholl, 2007). If this is the case, the stretch and challenge policy could (even assuming great examiner skill in devising questions that stretch and challenge) be well aimed at the wrong target. To know whether stretch and challenge has achieved its policy objectives, it would be important to monitor the impact of the new examinations upon teaching and learning following their introduction, once teachers have experienced the new examinations and had the time to adapt their approaches.

We would argue that the main purpose of A-levels should be to assess students' knowledge and understanding of the subject area, as defined by the syllabus. Using stretch and challenge in A-levels as a washback tool, or to produce measurement-driven-instruction could interfere with the main purpose of A-levels if the new-style questions produce de-motivated, anxious students or are in some other way unreliable or lacking validity. Thus, it will also be important to monitor whether, in achieving the stretch and challenge policy, the main purpose of A-levels is preserved unscathed.

7 Acknowledgements

The authors would like to thank the students who took part in these interviews, who engaged with the issues being put to them and were willing to be open. We would also like to thank Louise Lloyd who assisted with the data collection; Martyn Wright and Jackie Brown for their help in recruiting participants; Caroline Bowsher, Christine Rogers, Sal Allman and Charlotte Web for helping us to

contact students; and Professor Jannette Elwood for a critical review of a draft of this report and extending our knowledge of the literature. Funding for this work was provided by the Qualifications and Curriculum Authority, as part of their evaluation of the introduction of the stretch and challenge policy for A-level examinations. Teachers' views on the new A-level questions were investigated in a separate study.

8 Disclaimer

The views expressed in this report are those of the authors and are not necessarily endorsed by the Assessment and Qualifications Alliance, the University of Bristol or the Qualifications and Curriculum Authority.

9 References

- ACCAC (2003) A curriculum of opportunity: developing potential into performance. Meeting the needs of more able and talented pupils. Qualifications, Curriculum and Assessment Authority for Wales, Cardiff.
- ACCAC (2006) 'Meeting the Challenge' Quality Standards in Education for More Able and Talented Pupils. Qualifications, Curriculum and Assessment Authority for Wales, Cardiff.
- Ahmed, A. and Pollitt, A. (2001) *Improving the validity of contextualised questions*. Paper presented at the British Educational Research Association Annual Conference, Leeds, September. Accessed 9 March 2008.
http://www.cambridgeassessment.org.uk/ca/digitalAssets/113953_Improving_the_Validity_of_Contextualised_Questions.pdf
- Airasian, P.W. (1988) Measurement-driven instruction. A closer look. *Educational Measurement: Issues and Practice*. Winter, 6 – 11.
- Bailey, R., Pearce, G., Winstanley, C., Sutherland, M., Smith, C., Stack, N. and Dickenson, M. (2008) A systematic review of interventions aimed at improving the educational achievement of pupils identified as gifted and talented. Technical report. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Baird, J., Ebner, K., and Pinot De Moira, A. (2003) *Student choice of study in Curriculum 2000*. AQA Internal Report RC237.
- Baird, J. & Lee-Kelley, L. (2009) The dearth of managerialism in the implementation of national assessment policy. *Journal of Education Policy*, 24, (1), 55-81.
- BBC (2005) A-levels and GCSEs too easy. 25 August. Accessed 5th March 2009.
<http://news.bbc.co.uk/1/hi/england/oxfordshire/4182212.stm>
- Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2003) *Assessment for Learning. Putting it into practice*. Open University Press, Maidenhead, UK.
- Bourdieu, P. (1986) The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education*. 241 - 258. New York: Greenwood.
- CCEA (2006) Gifted and talented children in (and out of) the classroom: A report for the Council of Curriculum, Examinations and Assessment. February 28th. Belfast, Northern Ireland.
- CCEA (2007) Gifted and Talented Pupils: Guidelines for Teachers. Council of Curriculum, Examinations and Assessment: Belfast, Northern Ireland.
- Chapman, D.W. and Snyder, Jr. C.W. (2000) Can high stakes national testing improve instruction: reexamining conventional wisdom? *International Journal of Educational Development*, 20, 457 – 474.
- Cizek, G.J. and Bunch, M.B. (2007) *Standard setting. A guide to establishing and evaluating performance standards on tests*. Sage: California.
- College Entrance Examination Board (2009) *About AP* Downloaded from

-
- <http://www.collegeboard.com/student/testing/ap/about.html> on 23rd January 2009.
- Crisp, V., Sweiry, E., Ahmed, A. and Pollitt, A. (2008) Tales of the unexpected: the influence of students' expectations on question validity and implications for writing exam questions. *Educational Research*, 50 (1), 95-115.
- Curnock-Cook, M. (2005) Letter to Ruth Kelly on A-level stretch and challenge. 26 October. Accessed 7 September 2007. http://www.qca.org.uk/qca_8907.aspx.
- Daly, A. and Pinot de Moira, A. (2008) *Students' approaches to learning and their performance in the extended project pilot*. Assessment and Qualifications Alliance Unpublished Report RPA_08_AD_RP_057.
- Damiani, V.B. (2004) *Portfolio assessment in the classroom*. Bethesda MD: National Association of School Psychologists
- DCSF (2007) Effective Provision for Gifted and Talented Students in Secondary Education. Department for Children, Schools and Families: London. Accessed 30th October 2008. <http://publications.teachernet.gov.uk/default.aspx?PageFunction=productdetails&PageMode=publications&ProductId=DCSF-00830-2007&>
- DCSF (2008a) Handbook for Leading Teachers for Gifted and Talented Education. Department for Children, Schools and Families: London. Accessed 30th October 2008. <http://publications.teachernet.gov.uk/default.aspx?PageFunction=productdetails&PageMode=publications&ProductId=DCSF-00577-2008&>
- DCSF (2008b) Identifying Gifted and Talented Learners: Getting Started. Department for Children, Schools and Families: London. Downloaded from http://ygt.dcsf.gov.uk/FileLinks/312_new_guidance.pdf on 30th October 2008
- Department for Education and Skills (2005) *14-19 Education and Skills*. Accessed 1 October 2008: <http://publications.dcsf.gov.uk/default.aspx?PageFunction=productdetails&PageMode=publications&ProductId=CM%25206476>
- Dhillon, D. (2003) *Predictive models of question difficulty: A critical review of the literature*. AQA Research Paper, RC209.
- Dweck, C. (2006) *Mindset. The new psychology of success*. Ballantine books, New York.
- Gilbert, C. (2008) *Ofsted's response to the Third Report from the Children, Schools and Families Committee, Session 2007-08*. Select Committee on Children, Schools and Families Fifth Special Report. Appendix 2. <http://www.publications.parliament.uk/pa/cm200708/cmselect/cmchilsch/1003/100305.htm>.
- Gordon S, Reese M (1997) High stakes testing: worth the price? *Journal of School Leadership* 7, 345-368.
- Harlen, Wynne and Crick, Ruth Deakin (2003) 'Testing and Motivation for Learning', *Assessment in Education: Principles, Policy & Practice*, 10, 2, 169 — 207.
- Hayward, G. and McNicholl, J. (2007) Modular mayhem? A case study of the development of the A-level science curriculum in England. *Assessment in Education: Principles, Policy and Practice*, 14, 3, November, 335 - 351.
- Hill, C. (2004) Failing to meet the standards: The English Language Arts Test for fourth graders in New York State. *Teachers College Record*, 106, 1086-1123.
- Hodgson, A. and Spours K. (2003) *Beyond A-levels*. Kogan Page: London.
- Impara, J. C., & Plake, B. S. (1998) Teachers' ability to estimate item difficulty: A test of the assumptions in the Angoff standard setting method. *Journal of Educational Measurement*, 35, 69-81.
- International Baccalaureate Organisation (IBO) (2009) *Diploma Programme Core Requirements* Downloaded from <http://www.ibo.org/diploma/curriculum/core/essay/> on 23rd January 2009.
- Ipsos MORI (2009) *Perceptions of A-levels and GCSEs Wave 7*. January. Social Research Institute.
- Jakwerth, P. R., Stancavage, F. B., & Reed, E. D. (1999) An investigation of why students do not respond to questions. (A NAEP Validity Studies Panel Paper) Washington, DC: American Institutes for Research.
- James, M. (2000) Measured lives: the rise of assessment as the engine of change in English Schools. *The Curriculum Journal*, 11, 3, 343-364.
- Johnsen, S.K. (2007) *Portfolio Assessment of Gifted Students in Alternative Assessments for Identifying Gifted and Talented Students* edited by VanTassel-Baska, J.L. Prufrock Press Inc., Austin, TX, USA
-

-
- Kelly, R. (2006) Letter to Ken Boston on A-level stretch and challenge. 5 February. http://www.qca.org.uk/qca_8714.aspx. Accessed 7 September 2007.
- Kline, P. (2000) *Handbook of psychological testing*. Second edition. Routledge, London.
- Little, C.A. and Ellis, W.T (1991) *Aligning Curricula for the gifted with content standards and exemplary secondary programs* in *Content-Based Curriculum for High-Ability Learners* edited by VanTassel-Baska, J. and Little, C.A. Prufrock Press Inc., Austin, TX, USA
- Maxwell, J.A. (2005) *Qualitative Research Design: An Interactive Approach*. Applied Social Research Methods Series, volume 41. Sage, London.
- McNeil, L. and A. Valenzuela. 2001. "The Harmful Impact of the TAAS System of Testing in Texas: Beneath the Accountability Rhetoric," pp.127-150 in M. Kornhaber and G. Orfield, eds., *Raising Standards or Raising Barriers? Inequality and High Stakes Testing in Public Education*. New York: Century Foundation, 127-150.
- Messick, S. (1996) Validity and washback in Language Testing. *Language Testing*, 13, 3, 241-256.
- Meyer, L. (2003) Repeat of AQA Standards Unit analyses originally run for the GCE Economics awarding meeting, June 2003. Internal unpublished AQA Standards Unit Paper.
- Mouchel Parkman (2007a), *Institutional Quality Standards in Gifted and Talented Education: User Guide*, Downloaded from <http://ygt.dcsf.gov.uk/LibraryResources.aspx?libraryId=12> on 5th November 2008.
- Mouchel Parkman (2007b) *Classroom Quality Standards: User Guide*. Downloaded from <http://ygt.dcsf.gov.uk/LibraryResources.aspx?libraryId=12> on 13th November 2008.
- Murphy, R. & Warmington, P.C. (2002) *Student perceptions of the new A-Level examinations and media reports on published results*. Joint Council for General Qualifications, The University of Birmingham, School of Education.
- Murphy, R., Greatbatch, D., Wilmut, J., Tolley, H., Mallia, C. and Hamer, J. (2007) *Improving Written GCSE Examinations Project: Interim Report to QCA from the Centre for Developing and Evaluating Lifelong Learning*. School of Education, Nottingham: CDELL.
- OCR (2007) *An evaluation of the 2007 stretch and challenge trials*. A study conducted on behalf of QCA.
- Office of the Qualifications and Examinations Regulator (2008) *Predictability studies report: A study of GCSE and GCE level examinations*. Ofqual/08/3866.
- Paris, N. (2007) Winchester drops too easy A-levels. *The Telegraph*, 7 July. Accessed 5th March, 2009. <http://www.telegraph.co.uk/news/uknews/1556713/Winchester-drops-too-easy-A-levels.html>.
- Paton, G. (2009) 'Bright pupils agonising over simple questions' says Eton Head. Accessed 5th March, 2009. <http://www.telegraph.co.uk/education/educationnews/4938107/Bright-pupils-agonising-over-simple-questions-says-Eton-head.html>.
- Pintrich, P. R., & De Groot, E. V. (1990) Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33–40.
- Pollard, A., Triggs, P., Broadfoot, P., McNess, E. and Osborn, M. (2000) *What pupils say: changing policy and practice in primary education*. Continuum, London.
- Pollitt, A. and Ahmed, A. (1999, May) *A new model of the question answering process*. Paper presented at International Association for Educational Assessment Annual Conference, Slovenia. Accessed 15 May 2008: http://www.cambridgeassessment.org.uk/ca/digitalAssets/113788_A_New_Model_of_the_Question_Answering_Process.pdf
- Pollitt, A. and Ahmed, A. (2000) *Comprehension failures in Educational Assessment*. Paper presented to European Conference on Educational Research, Edinburgh, September.
- Pollitt, A., Ahmed, A., Baird, J., Tognolini, J. and Davidson, M. (2007a) *Improving the Quality of GCSE Assessment. Interim Report*. London: Qualifications and Curriculum Authority.
- Pollitt, A., Ahmed, A. and Crisp, V. (2007b) *The demands of examination syllabuses and question papers*. In Newton, P., Baird, J., Goldstein, H. Patrick, H. and Tymms, P. (Eds.) (2007) *Techniques for monitoring the comparability of examination standards*. London: Qualifications and Curriculum Authority.
- Pollitt, A., Hutchinson, C., Entwistle, N. and de Luca, C. (1985) *What makes exam questions difficult? An analysis of 'Grade' Questions and Answers*. Edinburgh: Scottish University Press.
- Popham, W. J. (1987) The merits of measurement-driven instruction. *Phi Delta Kappan*, 68, 679-682.
-

-
- Priestley, Mark (2003) 'Curriculum 2000: a broader view of A-levels?', *Cambridge Journal of Education*, 33:2, 237-255.
- Reay, D & Wiliam, D. (1999) 'I'll be a nothing': structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25, 343-354.
- Renzulli, J.S. and Callahan, C.M. (2007) *Product Assessment in Alternative Assessments for Identifying Gifted and Talented Students* edited by VanTassel-Baska, J.L. Prufrock Press Inc., Austin, TX, USA.
- Sadler, D.R. (2007) 'Perils in the meticulous specification of goals and assessment criteria'. *Assessment in Education: Principles, Policy & Practice*, 14, 3, 387-392.
- Sousa, D.A. (2003) *How the Gifted Brain Learns* Corwin Press. Thousand Oaks, CA, USA.
- Sundre, D.L. (1999) *Does examinee motivation moderate the relationship between test consequences and test performance?* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada. April 19 – 23.
- Sundre, D.L. and Kitsantas, A. (2004) An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology*, 29, 1, January, 6 – 26.
- VanTassel-Baska, J. (2000) *Curriculum Policy Development for Secondary Gifted Programs: A perception for reform coherence* National Association of Secondary School Principals Bulletin April vol 84.
- VanTassel-Baska, J. (2007) *Alternative assessments for identifying gifted and talented students*. Prufock Press Inc., Austin, Texas.
- VanTassel-Baska, J. and Little, C.A. (1991) *Content-Based Curriculum for High-Ability Learners* National Association for Gifted Children. Storrs, CT.
- Wiliam, D. (2001) What is wrong with our educational assessments and what can be done about it. *Education Review*, 15, 1, 57-62.
- Wiliam, D. (2007) *Comparative analysis of assessment practice and progress in the UK and USA*. Paper presented at Westminster Education Forum Seminar on Assessment, October, London.
- WJEC (2007) GCE Music stretch and challenge: a study into the effect of applying new criteria. A study conducted on behalf of QCA.
- Wolf, A. (2002) *Does education matter? Myths about education and economic growth*. Penguin books.
- Wolf, Lisa F., Smith, Jeffrey K. and Birnbaum, Marilyn E. (1995) 'Consequence of Performance, Test, Motivation, and Mentally Taxing Items', *Applied Measurement in Education*, 8:4, 341 — 351
- Zeidner, M. (1993) Essay versus multiple-choice type classroom exams: The student's perspective. In B. Nevo, & R. S. Jager (Eds.), *Educational and psychological testing: The test-takers' outlook* (pp. 67–82) Toronto, Ontario, Canada: Hogrefe & Huber.

STUDENTS' VIEWS OF A-LEVEL [PSYCHOLOGY/BIOLOGY]

Did you take an A-level in [Psychology/Biology] last summer?

Would you like to earn a £40 book token for less than an hour's work?

If so, we would like to hear from you. The AQA & University of Bristol are looking for 12 undergraduates to take part in a research project exploring students' experiences of A-level exams - whether you found them easy or difficult and whether you enjoyed them or not – your opinions and experiences are important to us.

We would like to interview students who sat an A-level [Psychology/Biology] exam last summer for no more than 1 hour each. It doesn't matter which exam board you took your exam with – all welcome.

If you would like to be considered for this project, please click on the following link and enter your details. If you are selected to take part in the study we will be in touch within the next couple of weeks to arrange an appointment. The interviews will take place in the week beginning [1st/8th] December.

http://www.surveymonkey.com/s.aspx?sm=UMNJ2gpA_2bByCJm2Z2IbUig_3d_3d

We look forward to hearing from you!

Appendix B Interview schedule notes

Introduction (2-3 mins)

Confirm who the interviewee is.

Introduce self.

Introduce project and conditions:

- This interview is part of a research study into students' experiences of sitting A-level examinations. It is jointly conducted by the University of Bristol and the exam board AQA. The Qualifications and Curriculum Authority have provided the funds for it.
- All information you give us will be held in confidence and it will be destroyed at the end of the study. Any information you give us will be reported anonymously so that it will not be possible to identify you in the reports.
- When the final report is complete, we will send you a one page summary of the findings.
- We will record this interview and the recording will be transcribed. I will also make some hand-written notes during the interview. The recording, transcription and notes will be destroyed once we have analysed the data. We will also destroy records of your contact details.
- If you wish to withdraw from the study, then please say so now. The consequence will be you will not receive the book token.
- Please sign the consent form to say that you agree for us to use the information you provide.
- And here is a sheet of information for you to take away.

Now we will proceed with the interview. Please take your time to answer the questions. There are no right or wrong answers, we are simply interested to hear what you have to say.

We will be talking about your experiences of taking Biology / Psychology A-level in the summer.

(Have past and sample papers ready.)

Section 1: EXPERIENCES OF GCE (time)

Which is the paper you took for your A2?

CONCEPTS	PROMPTS
Recall	Could you tell me what you remember about sitting this paper? (<i>Do you remember any of the questions? What made them stand out? How did you feel on the day? Nervous, calm, anxious, confident?</i>) What grade did you get for this paper? Perhaps you know your UMS score.
Motivation including self-efficacy	How important was it that you got this grade / score? (<i>What was at stake when you took this paper? Was university entrance resting on it? What motivated you to study for this paper? Competition with classmates/siblings? Pleasing your parents/teachers? Proving something to yourself?</i>) Did you get the grade you were expecting? (<i>What were you expecting?</i>)

	<p><i>Either: You thought you would get an X but you got a Y – why did you think you would get an X? Do you think you had a good idea of your ability or was there something about the exam paper?</i></p> <p><i>Or: You were expecting an X and you got an X - why did you think you would get an X? Do you think you had a good idea of your ability or do you think the exam paper was predictable?)</i></p>
Revision or study strategies	<p>Please tell me something about how you prepared for this paper? <i>(What were your study strategies or your revision strategies? How many hours? Learning by heart, question practice, revision timetable, targeting content areas.)</i></p> <p>How prepared did you feel for it? <i>(Confident, good on some topics not on others, anxious)</i></p> <p>Did your teachers make special efforts to help you? What did they do? <i>(Practice questions, advice on exam strategy, leave it up to you, sample answers)</i></p> <p>With hindsight, what do you think was the best strategy for excelling in this paper? <i>(Learning off by heart, practice, target particular questions, writing out answers again and again)</i></p>
Difficulty perception	<p>I'd like you to tell me how easy or difficult this paper was. Can you rate it out of 10, with 1 = very easy and 10 = very difficult.</p> <p>What made it easy/difficult?</p> <p><i>(Language, layout, content, preparation, use of practice papers, type of questions, structure of questions, teacher prepared you for it;</i></p> <p><i>Met expectation in terms of difficulty? If not, why was this? Didn't revise, wasn't an important paper for overall grade so didn't focus on it, questions came up you weren't prepared for.</i></p> <p><i>If yes, why was this – preparation, past papers, practice..)</i></p> <p>Did you feel challenged by this paper? Why do you say that?</p> <p>Going back to what you said about how you prepared for your exams, did you prepare differently for papers that you thought would be easy/hard?</p>
Stimulating	<p>Did you find this exam paper interesting? <i>(Did it engage you in some way other than for testing you? In what way?)</i></p>

Section 2: WHAT DOES STRETCH AND CHALLENGE ADD? (time)

This is a sample paper for 2010 of the paper you took in 2008. It is accessible from AB's website.

I am interested to know what you think of the questions in this sample paper. Please read it alongside the paper you sat.

Identifying stretch and challenge	<p>How do these two papers compare? Can you see any differences between them? What?</p> <p><i>Is one more challenging than the other? What makes you say that? Is it the style of question, content, task, language?</i></p> <p><i>Responses related to S&C initiatives: less formulaic, avoid atomistic Qs, fuller coverage, require paragraph answers, wider range of item types, use of stimulus material such as case studies, not relying on recall)</i></p> <p>Is this paper easier, more difficult or the same as the one you took? What rating would you give it with 1=very easy, 10=very difficult? Why do you say that? What makes it more/less difficult?</p>
Strategies for answering stretch and challenge questions	<p>How would you feel if you had to take this paper instead of the one you took? <i>More or less confident? What would you do differently if you were preparing to take this paper? Would you change your study strategies? How?</i></p> <p>Would your teachers need to do anything differently to prepare you? What? <i>More help? Target analytical thinking? More practice? Master-classes, more resources, model answers?</i></p> <p>Do you think you would feel more or less challenged by this paper? Is that a good or bad thing? Why do you say that?</p>

Section 3: GENERAL PERCEPTIONS OF STANDARDS (time)

Perception of exam difficulty	<p>In your opinion, what makes an exam question difficult? (<i>Content, style of question, language, lack of preparation</i>)</p>
Perception of national standards	<p>Have you heard people and the press say that A-levels are becoming too easy and everyone passes now? In your opinion, is this true? What would you say to them? Do you think they should be challenging? More challenging than they are now? Why?</p> <p><i>(Separate out abilities, push students to work harder, make the jump to HE easier,)</i></p>

Any questions? Anything else to add? Thanks, book token.



CONSENT FORM

A-level Research Project

I(Name)

agree to take part in the above research project. I give my permission for any data I supply to be recorded and transcribed for data analysis purposes. I understand all data collected will be anonymised and any resultant reports will not identify specific individuals. I also understand I may withdraw from the research at any time without giving a reason for doing so.

Signed (participant)..... Date.....

Signed (researcher)..... Date.....



A-LEVEL RESEARCH PROJECT

Thank-you for taking part in this research project. The research is being conducted by the Assessment & Qualification Alliance (AQA) and the University of Bristol, and is funded by the Qualifications & Curriculum Authority (QCA). Please find below further information on the background to the research, and the project itself.

Background

In recent years, it has been argued that A-levels do not stretch the most able students. In response to these concerns, the Government introduced a policy that has become known as 'Stretch and Challenge'. New A-level examinations have been introduced, for first teaching in September 2008, with final year examinations that are intended to include new questions designed to 'stretch and challenge' students. The intention behind the introduction of stretch and challenge questions is to affect students' experiences of sitting the examinations. Highly able students are intended to experience the examinations as taxing their skills, rather than being well within their capabilities

Research project

This project has been designed to explore students' perceptions of A-level examinations and the new 'Stretch & Challenge' style A-level questions. As part of the project interviews are being conducted with students from Bristol and Manchester University, who sat an A-level in Psychology or Biology in June 2008. All interview data is being recorded and transcribed for analysis purposes; the analysis will involve the researchers searching for key themes in the data and coding these. All responses will remain confidential, however, and will be anonymised in any reports.

The consent form that you signed at the beginning of your interview gave your written permission for the interview to be recorded and transcribed. On completion of the study the audio recordings and all personal data will be destroyed. You do, however, have the right to withdraw from the study at any point. Should you choose to withdraw from the study your interview recording will be deleted and any transcription destroyed.

A report of the findings from this project will be produced and submitted to QCA; this will form part of an evaluation of the 'Stretch & Challenge' initiative. When the final report is complete we will send you a summary of the findings. The final report will also be posted on the QCA website. The results may also be used for conference presentations, and may form part of a journal article that will be submitted to an academic journal.

Thank-you again for taking part in this project – we hope you found it an interesting experience. Your experiences and perceptions are an important part of this research.

If you have any queries or complaints following your interview today please contact Martyn Wright or Jackie Brown on [telephone number given].

Appendix E Pedagogical approaches to exam preparation

Example study techniques

- A teacher who set up a 'star system'; an analysis of past questions and the probability of each question appearing that year, with probability indicated by the number of stars;
- Producing as many as three revision booklets for each *question*;
- A CD-ROM of all the available past papers and mark schemes;
- A booklet of model answers for every question appearing in recent years;
- A group song and dance routine to aid revision;
- Team-teaching or co-teaching alongside the teacher;
- Group debates and subsequent sharing of 'for' and 'against' notes;
- Inviting guest speakers from local business and parents with expertise;
- Re-enacting research studies in biology and psychology;
- Setting up a crime scene puzzle to be solved using revision material;
- Revision podcasts;
- Students marking the practice past paper responses of other students using the mark scheme.

Students were taught the following examination techniques:

- How to read and deconstruct questions (eg break into parts, underline key words);
- Understand what they were being asked to do (understand the meaning of each command word and what this meant for the way they were required to respond to the question);
- How to manage time in an exam and, in particular, making sure they appropriately allocated minutes to marks;
- The difference between assessment objectives - AO1 (description) and AO2 (evaluation), including being able to identify whether a question was asking for AO1 or AO1 and AO2, and how to structure their responses to access the AO1 and AO2 marks;
- How to ensure the award of marks by using (the correct number of) point-scoring key words, phrases or arguments.