

Review of Standards in GCE A level Critical Thinking

2010



April 2012

Ofqual/12/5158

Contents

Executive summary	3
Findings	3
Section 1: Introduction	4
Context.....	4
Methodology of the review	4
Provision of assessment materials and student work	5
The review team.....	6
Analysis of the specifications and assessment materials	6
Analysis of student performance	6
Data analysis.....	7
Section 2: Subject demand in GCE A level critical thinking	8
Overview	8
Findings	8
Assessment objectives	8
Specification content	9
Schemes of assessment	9
Question papers and mark schemes	11
Section 3: Standards of performance	15
Overview	15
Process	15
Interpreting the graphs	15
Findings	15
Performance at the GCE AS grade-A boundary in 2010	16

Performance at the GCE A2 grade-A boundary in 2010.....	16
Performance at the GCE AS grade-E boundary in 2010	17
Performance at the GCE A2 grade-E boundary in 2010.....	17
Conclusions.....	19
Appendix A: Provision of assessment materials and student work at GCSE and GCE levels for Ofqual's archive (annual inclusion and standards reviews).....	20
Section 1: Specification of requirements	20
Section 2: Student work	21
Appendix B: Availability of specification materials for the purposes of this review....	23
Appendix C: Student achievement by grade.....	24
Appendix E: Measure, standard error (SE) and infit values of the ranked scripts.....	25
Appendix F: Review team.....	30

Executive summary

The Office of Qualifications and Examinations Regulation (Ofqual) undertakes a rolling programme of reviews across high-profile GCSE and GCE A level subjects to monitor whether standards in assessment and student performance have been maintained over time.

This report details the findings for GCE A level critical thinking in the year 2010. This is the first time we or our predecessors have reviewed standards in this subject.

The review compared subject specifications, assessment materials and student work from the two awarding organisations awarding this qualification in the year 2010 being reviewed, AQA and OCR.

Findings

We found that the content of the qualifications and the progression between AS and A2 differed between the awarding organisations. For example, the AQA specification contained more challenging topics and more technical vocabulary at AS making it more demanding. However, at A2, question papers were similar in style to their AS question papers, limiting the amount of progression between the two levels. This contrasts with OCR, where while the specification was less demanding at AS than AQA's, it included more open-ended questions at A2, thus offering students greater opportunity to demonstrate higher levels of critical thinking, including synthesis.

Section 1: Introduction

Context

We regularly review qualifications in different years to check that standards are maintained over time.

These reviews inform future developments in qualification and subject criteria and help us to compare standards across awarding organisations. In our reviews we:

- analyse the nature of the requirements different assessments make on students
- compare the levels of performance required for a particular grade in different assessments
- consider how these two elements relate to each other.

GCE A level critical thinking is offered only by AQA (specification codes 1771/2771) and OCR (specification codes H052/H452). The number of students taking GCE A level critical thinking was 1,878 in 2011 with 18,819 students taking the qualification at AS only. A detailed breakdown of student-entry numbers and cumulative percentage pass rates can be found in [Appendix C](#).

This is the first review of standards in GCE A level Critical Thinking. Prior to the new A levels (first teaching 2008) critical thinking was offered only by OCR. GCE A level critical thinking is an unusual subject in that there is not a significant body of subject knowledge which students are expected to learn. Instead the focus is on the acquisition of a range of skills which support analysis, evaluation and development of reasoned argument.

Methodology of the review

Standards reviews examine different specifications within a qualification, the associated assessment instruments and samples of student work, by collating and analysing the views of a number of subject specialists. The following sections of this report detail the process of collecting and processing this information. In these reviews, we compare how demanding a specification is against all of the other specifications under review and includes consideration of:

- specification-level factors such as assessment objectives, content and structure
- assessment-level factors such as what content is assessed, the weighting of each component and how the assessments are marked
- student performance-level factors, including how the students responded to the assessments and the grades they received as a result.

How demanding an assessment of a qualification is can be defined in a variety of ways and is linked to the purpose of the qualification. It is related to:

- the amount and type of subject knowledge required to be assimilated
- the complexity or number of processes required of the students, the extent to which the students have to generate responses to questions from their own knowledge, or the extent to which resources are provided
- the level of abstract thinking involved
- the extent to which the students must devise a strategy for responding to the questions.

Provision of assessment materials and student work

Details of our requirements for the provision of assessment materials and student work for review are given in [Appendix A](#) and, in summary, include:

- the current specification
- all associated question papers
- final mark schemes
- reports from the examiners and grade boundaries (overall and by unit, and both raw and scaled)
- mark distributions, grade descriptors and assessment grids
- any other information that was routinely supplied to centres
- all the assessment work carried out by a sample of students whose final grade lay at or near the judgemental grade boundaries for the qualification being analysed.

No equivalent materials had been collected and retained previously for this qualification.

Full details of the materials supplied by awarding organisations can be found in [Appendix B](#).

The review team

We contracted six experts in GCE A level critical thinking to undertake the review. These reviewers were sourced through:

- a subject-expert recruitment exercise carried out by us in November 2010, advertised via the Times Educational Supplement and our website and newsletter
- nominations made by the regulators in Wales and Northern Ireland
- nominations made by awarding organisations involved in the review
- nominations made by subject associations and other learned organisations invited to participate in the review.

A full list of reviewers can be found in [Appendix F](#).

We contracted a lead reviewer, specification reviewers and script reviewers. (All nominees from awarding organisations and subject associations were script reviewers.)

Analysis of the specifications and assessment materials

The lead reviewer and specification reviewers (specification review team) analysed the awarding organisations' materials, using a series of forms which can be found via the comparability page on our website at www.ofqual.gov.uk/research-and-statistics/research-reports/92-articles/23-comparability .

These analyses are designed to describe how demanding the specification is.

Each reviewer analysed a subset of the specifications available, so that there were at least three different views on each specification. The lead reviewer then produced a report which brought together the views of the reviewers on each of the awarding organisations' specifications. The specification review team was given the opportunity to discuss the lead reviewer's conclusions at a follow-up meeting. These findings are presented in Section 2 of this report.

Analysis of student performance

To assess student performance, all reviewers were brought together for a two-day meeting to analyse students' scripts (pieces of student work supplied by the awarding organisations). This process is referred to as a script review. The meeting started with a briefing session to make sure that all the reviewers had a common understanding of the methodology and the judgement criteria.

The scripts were organised into packs for consideration during the review. Packs were organised by grade. (Only grade boundaries A/B and E/U were analysed, as grades B, C and D are calculated arithmetically after grade-boundary marks for grades A and E have been set during the awarding process carried out by the awarding organisations.)

As far as was possible, given the collection of scripts available, packs contained 12 scripts at the same grade, with at least two scripts from each awarding organisation from 2010 (the remaining two scripts were selected at random).

Reviewers were then asked to rank the 12 scripts in each pack, from best to worst, on a data-entry sheet and to make comments on the scripts as necessary. Each reviewer completed a maximum of 14 sessions over the two-day review.

During the two-day meeting there were plenary sessions for reviewers to discuss the script review process and the quality of the scripts being analysed.

Data analysis

We use a software package called FACETS to analyse the results from data-entry sheets produced during the script review. FACETS uses a Rasch model (often classified under item response theory) to convert the qualitative ranking decisions made by reviewers into a single list that reflects the probable overall order of the sets of student work, from best to worst.

We use this list, alongside the qualitative comments made during the script review and findings from the specification review, to inform Section 3 of this report.

Section 2: Subject demand in GCE A level critical thinking

Overview

Specification reviewers considered the amount and type of knowledge about critical thinking required by each awarding organisation's specifications. They did this by analysing specification documents, reports from the examiners and question papers with associated mark schemes from each of the awarding organisations for 2010. Details of the specification materials included in the review are given in [Appendix B](#).

The AQA specification was considered more challenging than the OCR specification. At both AS and A2, the AQA specification contains more challenging topics and more technical vocabulary, and has greater depth and breadth than that of OCR. Many of AQA's AS questions were worded similarly to OCR's and mark-scheme expectations were comparable. Significantly, student work provided for the review demonstrated very similar characteristics and levels of performance for the two qualifications at both AS and A2.

OCR's June 2010 question papers achieved better and more consistent coverage of specification content than AQA's. This was partly, although not entirely, due to the fact that AQA's specification had more content.

The two qualifications diverge in approach at A2. AQA's A2 question papers were similar in style to their AS question papers. However, OCR's A2 question papers included more open-ended questions, offering students greater opportunity to demonstrate higher levels of critical thinking, including synthesis.

Findings

Assessment objectives

For both qualifications, the assessment objectives and the weightings set out in the relevant specifications are in line with the GCE A level critical thinking subject criteria.

- Assessment objective 1 – analyse critically the use of different kinds of reasoning in a wide range of contexts.
- Assessment objective 2 – evaluate critically the use of different kinds of reasoning in a wide range of contexts.
- Assessment objective 3 – develop and communicate relevant and coherent arguments clearly and accurately in a concise and logical manner.

The assessment objectives suggest a hierarchy of skills, with assessment objective 3 a higher level skill than assessment objectives 1 and 2. However, in practice, question content and design also contribute to determining whether assessment of a particular objective is more or less demanding.

Specification content

The AQA specification was considered to be well designed, consistent and thorough in its topic coverage and should provide a solid foundation for academic study in a range of disciplines. The OCR specification provides a stronger sense that AS and A2 and the four units are each distinctive. The OCR specification topics are clearly defined and focused tightly on skills in terms of the assessment objectives. The wording of the topics is appropriate and is matched by the command words in the question papers, providing clarity and focus on aspects of teaching. This may, however, lead to opportunities offered by the specification (for exploration of real-life contexts through analysis, evaluation and argumentation) being overlooked in favour of training for the examination.

There is greater breadth and depth of content in AQA's specification than in OCR's. AQA specifies skills at AS which in the OCR specification appear at A2. For example, AQA's Unit 1 section 3.1.4a introduces "reasons acting independently or in combination (jointly)"; OCR does not expect understanding of this concept until Unit 4 section 3.4.1.

In both specifications, there was duplication of topics across units. This is to be expected. Critical thinking is skills-based with progression achieved through later units building on the content of earlier units as well as through the greater complexity of source material.

Schemes of assessment

The AQA and OCR qualifications have similar schemes of assessment, as shown in the tables on the next page.

AQA GCE A level Critical Thinking				
Unit/component title	Weighting (%)	Time allowance	Total marks	Type of assessment
Unit 1 – CRIT1 Critical thinking foundation unit	25	1 hour 30 minutes	70	Written exam
Unit 2 – CRIT2 Information, inference and explanation	25	1 hour 30 minutes	70	Written exam
Unit 3 – CRIT3 Beliefs, claims and arguments	25	1 hour 30 minutes	70	Written exam
Unit 4 – CRIT4 Reasoning and decision making	25	1 hour 30 minutes	70	Written exam

OCR GCE A level Critical Thinking				
Unit/component title	Weighting (%)	Time allowance	Total marks	Type of assessment
Unit 1 – F501 Introduction to critical thinking	25	1 hour 30 minutes	75	Written exam
Unit 2 – F502 Assessing and developing argument	25	1 hour 30 minutes	75	Written exam
Unit 3 – F503 Ethical reasoning and decision making	25	1 hour 30 minutes	60	Written exam
Unit 4 – F504 Critical reasoning	25	1 hour 30 minutes	60	Written exam

AQA's four June 2010 question papers were very similar in style, structure and presentation at AS and A2. They contained varying combinations of short-answer questions, structured questions requiring extended responses and unstructured questions requiring extended responses.

OCR's assessment model is very different. Unit 1 comprised a highly structured series of short-answer questions and a scaffolded question requiring an extended response. Unit 2 contained 15 multiple-choice questions, short-answer questions and two structured questions requiring extended responses. The A2 question papers comprised a small number of short-answer questions and unstructured questions requiring extended responses.

Apart from AQA's Unit 4, all units of both specifications require students to answer questions based on source material, seen for the first time in the examination. AQA's Unit 4 was based on a lengthy set of pre-released documents and an additional document issued in the examination. For all four units, AQA students had considerably more resource material to deal with in the examinations than OCR students had.

Both qualifications achieve progression as much through the greater complexity of resource material as through the greater demand of the second AS and the A2 question papers.

Question papers and mark schemes

AQA

AQA's question papers and source-material booklets were well presented and well laid out, with a good variety of source documents. Source documents generally dealt with interesting topics, but there was a large quantity of source material, especially at A2.

Many questions followed good assessment practice. Higher-attaining students were offered ample challenge, but there was evidence that lower-achieving students had insufficient opportunity to demonstrate their skills. Across all units, the top end of the mark range was not used, and mean marks were low for Units 1 and 4, suggesting that there are particular problems with those question papers and mark schemes.

Across all question papers, the combination of lengthy source material and demanding questions required students to think under time pressure. The Unit 1 principal examiner reported that "... most students completed all questions set", implying that a proportion did not complete all the questions.

Reviewers were concerned that the number of lines provided for answers sometimes appeared arbitrary. For example, in Unit 1, two-mark questions were variously allocated three, four or six lines, and this did not seem to relate to what the mark scheme expected. As noted in the Unit 3 chief examiner's report, "there is an optimum as well as a minimum length for acceptable answers" and "the space provided is a guideline".

In all question papers there were few easily accessible questions and some AS questions went beyond the appropriate level of demand, particularly in Unit 1. For example, Unit 1 question 2 asked students to consider "how much support" the analogy provided, and question 4 asked for an "implied intermediate conclusion".

The reviewers thought that assessment would be more coherent if question papers were targeted on the specific unit content, and if crossover to topics covered in other units was reduced.

It was also judged that, where skills were repeated at A2 there seemed to be little progression, in terms of the questions being less structured and the depth of thinking and insight required. Questions 3 to 8 in Unit 4 provided 29 of 70 marks by testing skills from other units, but not in ways that required synthesis.

The proportion of marks achieved through short-answer questions was high:

- approximately 70 per cent in Unit 1
- 60 per cent in Unit 2
- 60 per cent in Unit 3
- 50 per cent in Unit 4.

There were few questions which were not short-answer or extended-answer. However, a number of questions could have justified higher mark allocations such as Unit 1 question 8 and Unit 2 questions 3, 4, 5(b), 6, 7(c), 8(a)(c). Across the four units, some longer questions assessing assessment objective 2 would be welcomed. They would allow demonstration of synthesis of the skills such as assessing overall how well a piece of reasoning works or weighing up strengths and weaknesses to come to an overall view.

For questions that were highly focused on a specific skill or exact factual response, mark schemes described accurately the type of response required. However, there were instances where the mark scheme appeared to reward responses that would not necessarily have been prompted by the question. For example, Unit 3 question 8 directed students towards "claims, assumptions and the reasoning", but the mark

scheme's top level descriptor required "insightful critical comment on the validity... of each of the steps in the argument".

The quality of written communication was assessed under assessment objective 3 through generic marking grids. Criteria for the quality of written communication for Unit 3 did not correspond with those for the other three units. And there were inconsistencies in the wording across the other units, including Unit 4 question 9 which appears to treat communication as a hurdle for achievement of the higher levels. Descriptors such as "developed" for a higher level of written communication are not particularly clear to centres.

OCR

The scheme of assessment (pages 21 to 22 of the specification) stated that the question papers, for the different units, will present various forms of material (images, diagrams, charts or statistical representations) as well as text. With the exception of Unit 1, the question papers did not adhere to the scheme of assessment. Aspects of the assessment were artificial, in contrast to the variety of forms of communication students encounter every day.

The consistent use of command words that focus on the assessment objectives is a dominant feature of this specification and would improve accessibility for lower-attaining students. However, an apparently straightforward question could require extreme precision in the answer: "state" questions often assessed ability to copy as much as to identify a conclusion, especially as the use of ellipsis was penalised. Some questions might work better if re-phrased, for example: "How well does this reasoning work?" rather than "Evaluate..."

A more consistent approach within and across units would have improved presentation of the assessment material, for example the background information in Units 1, 3 and 4. Questions requiring more than one response were displayed in different ways in Unit 1 (questions 1(b), 3, 6, 7, 9(a) and (b)) and Unit 2 question 22(a), and there were inconsistencies in the number of lines provided for answers worth similar marks. While these are relatively minor matters, clarity of presentation helps students respond appropriately.

In Unit 1 reviewers identified a number of issues with the assessment of the "credibility" specification content. The ability to assess credibility of sources is a valuable research and life skill; however the approach taken in the question paper and the mark scheme resulted in superficial and formulaic answers. Further, the overuse of scaffolding meant that the opportunities for students to demonstrate the high-level skills of evaluation and synthesis were restricted. For example, in question 5: "Suggest one reason – you must give only a reason" tested whether students

could phrase a sentence so that it was technically a reason, which was mechanistic and inconsistent with practice in Unit 2.

The specification requires students to “identify a wider range of argument components” in Unit 2 than are tested in Unit 1, but the multiple-choice questions (and some Section B questions) tended to cover much of the same ground as Unit 1: evidence, reasons, conclusions and assumptions. Some Unit 2 skills, including evaluating analogy and inference, were not tested at all.

At A2 most questions were open-ended. Mark schemes were thorough and included level descriptors, indicative content and examples, as well as topic breakdowns. The use of levels of response allowed students at both E and A grades to be appropriately rewarded.

Reviewers thought that in Unit 3 it might have been easier for students to handle the resource documents in the examination if they had been presented on fewer sides of paper. In question 3, success depends on students’ ability to think of choices and criteria; students are at a disadvantage if they are unable to do so.

In Unit 4 the less formulaic approach of question 2 – where there was no single right answer but marks were awarded for the analysis and justification – enabled students to demonstrate their skills. Question 3 was complex but without it the OCR A2 would not have been sufficiently demanding.

Mark schemes were very thorough, with: possible responses; exemplar answers; levels; generally useful rationales; and grids showing specification topic coverage. In some instances thoroughness tended towards being excessively prescriptive and advantaging formulaic responses.

Section 3: Standards of performance

Overview

Process

Reviewers considered student work from AQA and OCR awarding organisations in 2010. Reviewers viewed 15 examples of student work at each judgemental grade boundary (A and E) for both AS and A2.

Interpreting the graphs

The graphs below show the spread of the student work as produced by the FACETS software. The dot indicates the measure related to the relevant ranked script, and the error bar “whiskers” represent the standard error of measurement (SEM) to the corresponding measure. The difference between sequential measures demonstrates the strength of the difference in the ranking position. Large differences would illustrate that scripts were less close in terms of similarity of student performance than small differences. So there could be a larger difference in judged student performance between scripts ranked 1 and 2 than between 2 and 3 (the difference in student performance is not necessarily the same between ranked positions).

The SEM illustrates the level of confidence that the measure is accurate, the greater the SEM the smaller the confidence levels. Therefore, large whiskers mean that there is less confidence that the measure was accurate. The whiskers illustrate the level of confidence, with upper and lower points at which the measure could lie.

The FACETS software will usually produce a rank order, even when there is little difference between the quality of the student work considered in the review. This is due to the natural slight variability between students who get the same mark. In these cases the rank order would show a relatively even spread of student work from different awarding organisations throughout the rank order.

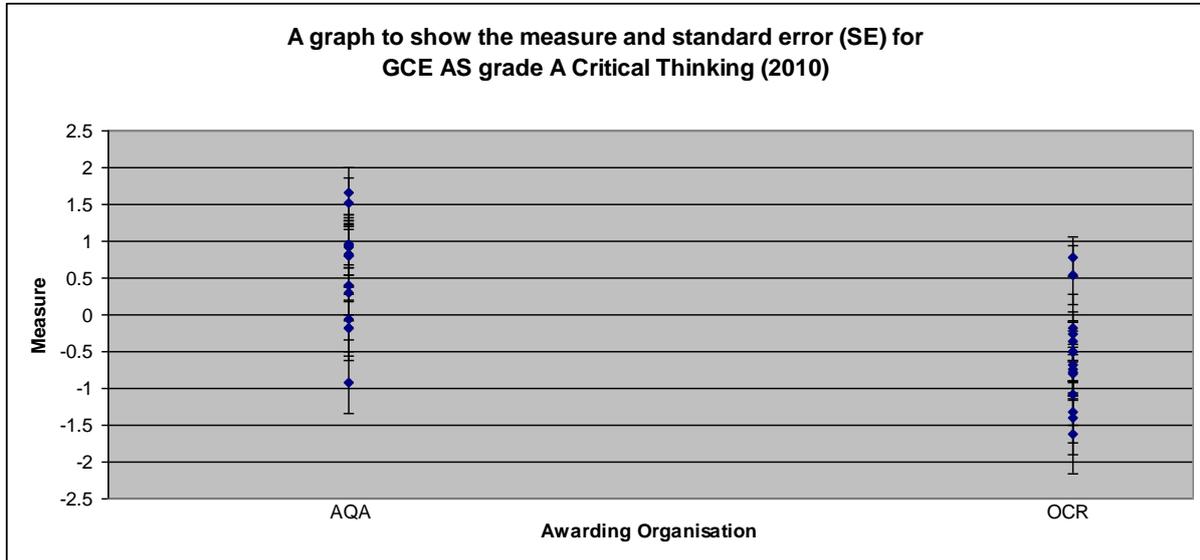
The scripts have been separated by awarding organisation for ease of reference, represented in alphabetical order across the horizontal axis (but can be found as a continuous inter-awarding organisation list in table format in [Appendix E](#)).

Findings

Overall, the student work was found to be comparable between the two awarding organisations, with work from AQA demonstrating a slightly higher standard of student performance.

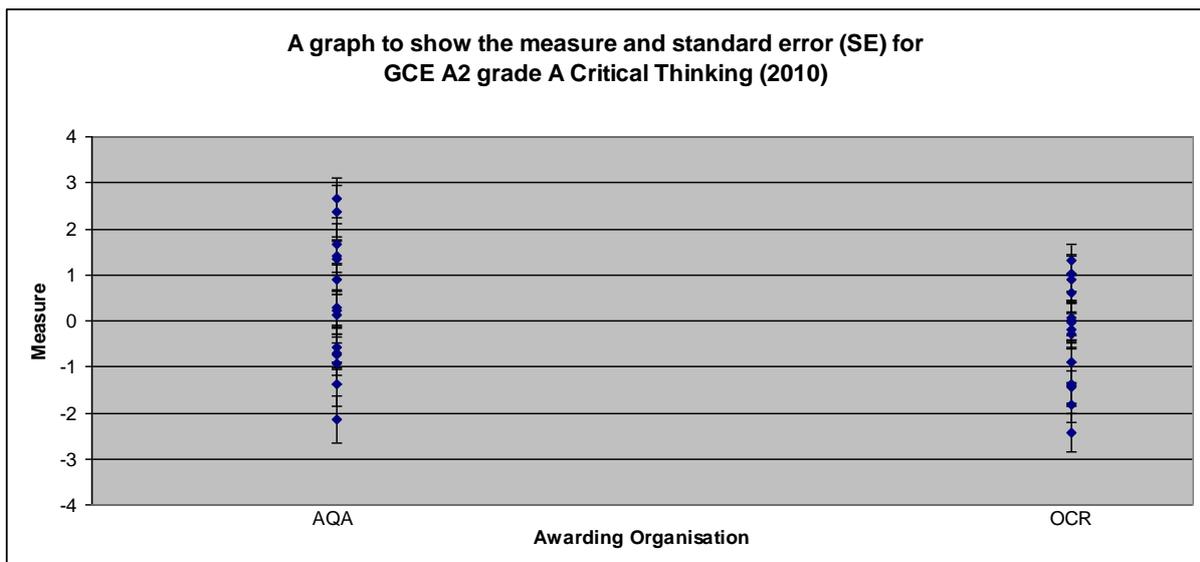
Performance at the GCE AS grade-A boundary in 2010

The majority of AQA’s student work was ranked within the top half of the ranking positions, suggesting a higher quality of student work at the grade boundary than that of OCR; only two of its scripts were in the lower ranked half.



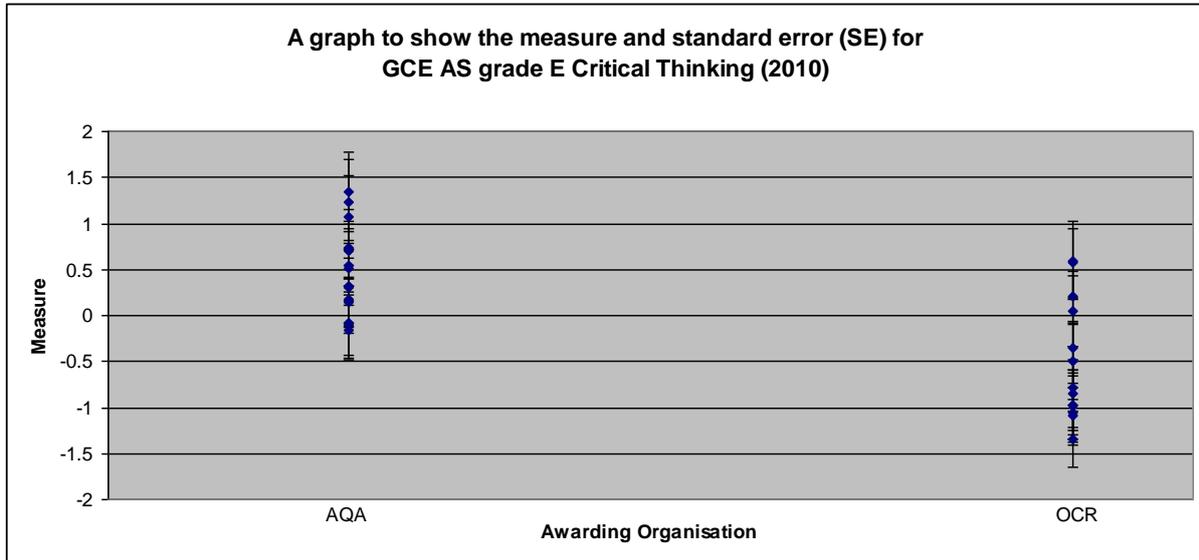
Performance at the GCE A2 grade-A boundary in 2010

AQA’s student work was ranked higher overall than OCR’s with more than half of AQA’s student work ranked within the top half of the ranking positions, suggesting a higher quality of student work at the grade boundary. However, the difference between the two awarding organisations was less marked than at AS for grade A.



Performance at the GCE AS grade-E boundary in 2010

The majority of AQA's student work was ranked within the top half of the ranking positions, suggesting a higher quality of student work at the grade boundary than that of OCR.

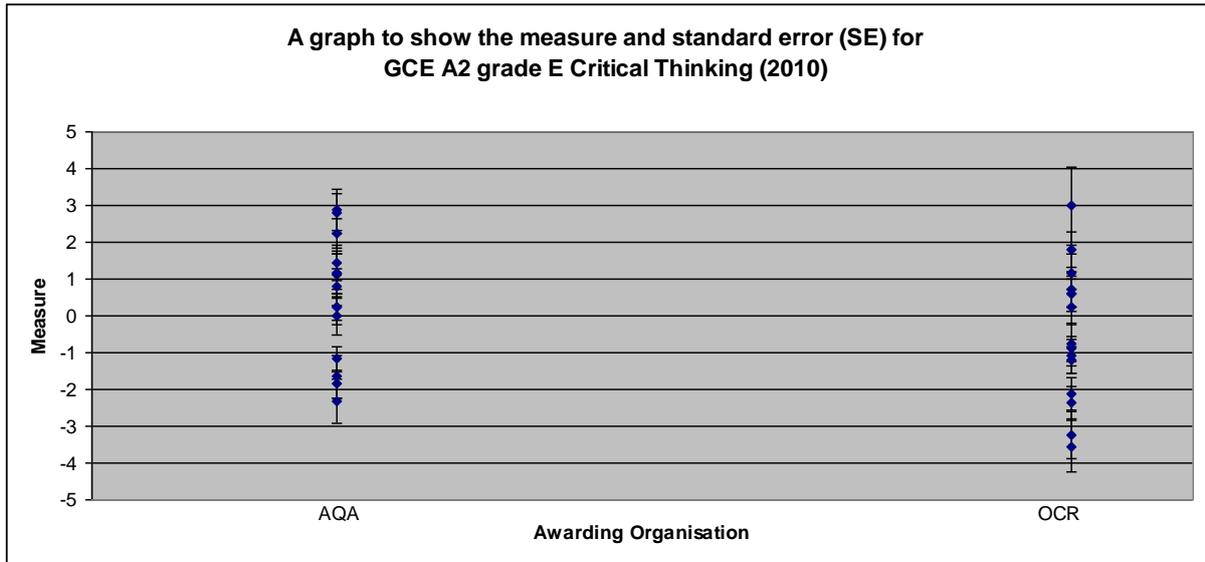


Performance at the GCE A2 grade-E boundary in 2010

AQA's student work was ranked higher than OCR's overall. However, the difference between the two awarding organisation was less marked than at AS for grade E.

More than half of AQA's student work was ranked within the top half of the ranking positions, suggesting a higher quality of student work at the grade boundary; six of its scripts were in the lower ranked half.

The scripts were, however, more frequently interspersed with one another at this grade when compared with grade A at AS and A2, which suggests a more consistent quality of student work between the two awarding organisations at this grade.



Conclusions

The AQA and OCR specifications and student work reviewed from 2010 indicated that the qualifications were sufficiently demanding at this level. However there were differences in the balance between AS and A2 units and the challenge and demand of question papers for both AS and A2 units.

We are talking to higher education institutions, amongst others, about the content and demand of A levels in the future, so that they meet the needs and expectations of people use this qualification in the future.

Changes to the qualification and their impact on perceptions of the academic demand of the qualification and its suitability as preparation for higher education will inform the next generation of qualification and subject criteria as part of this process.

Whether this variation in content and progression between specifications is acceptable in the future will form part of that discussion.

Appendix A: Provision of assessment materials and student work at GCSE and GCE levels for Ofqual's archive (annual inclusion and standards reviews)

Section 1: Specification of requirements

1.1 Each awarding organisation should draw the materials for each subject from the specification with their largest entry in summer 2009, unless that selection severely limits the range of examination components available. Where there are several entry options, materials should be drawn from the largest option only, unless Ofqual were exceptionally to agree other arrangements.

1.2 (With regards to GCSE) – where there are both modular and linear (non-modular) examinations in a subject, the awarding organisation operating the modular scheme with the greatest number of students (amongst all awarding organisations) should include that modular scheme, even if it is not a specification within the awarding organisation's largest entry. Similarly, the awarding organisation operating the linear scheme with the greatest number of students should include that linear scheme. If an awarding organisation runs both the largest entry linear examination and the largest entry modular examination in a subject, it will therefore provide two sets of materials, including student work, where required.

1.3 The following materials should be supplied:

a) Current specification: all associated question papers and final mark schemes.

b) The 2009 chief examiners' report (CER) and details of awarding procedures particular to the specification supplied.

c) An indication of how the specification's content and assessment criteria and objectives have been met in each question paper supplied. This may take the form of a grid. For objective tests this should include faculty values, discrimination indices and a specification grid detailing what grade each question was targeted at, as well as an indication of what percentage of students got a particular question correct when it was targeted at the grade they got overall.

d) Unit or component mark distributions (with grade boundary marks shown). It should be clear whether the marks are on the raw or uniform mark scale.

e) Grade boundaries, overall and by unit (both raw and scaled).

f) Student work as specified in Section 2.

g) Complete data record showing for each student selected the raw mark; final mark; weighted or uniform mark; grade for each component/unit (including any non-archived component/unit) and overall grade; and, where relevant, tier of entry.

Where appropriate, materials a)–e) may be supplied in electronic form.

Section 2: Student work

2.1 The work submitted should include the examination scripts, the internal assessment, and any oral/ aural examinations (with examiner mark sheet) where these are routinely recorded. In addition, for modular specifications, the examination papers of module tests should be supplied.

2.2 The sample should be of the original work of the students. Photocopies of work should only be used where it is impossible to send the originals and with agreement in advance by Ofqual. Student and centre names and numbers should be removed wherever they appear in a student's work, unless they form an integral part of the work, for example, within a letter.

2.3 Where an awarding organisation's specification has a relatively small entry or where, for some other reason, it is proving difficult to find sufficient students who fulfil the criteria, the awarding organisation should contact the Ofqual officer responsible to agree how best to finalise the sample.

2.4 All internal assessment submitted should be that of the particular students selected for the sample. If, for any reason, this proves to be impossible, the awarding organisation should contact the Ofqual officer responsible to agree appropriate alternative measures.

2.5 The sample of scripts retained for each specification (option) should be taken from students whose final mark lay at or near the subject grade boundaries for A/B, C/D and F/G for GCSE and A/B and E/U for GCE A level qualifications. At each boundary, each awarding organisation will supply the externally and internally set and marked assessments of fifteen students. Students selected should be those whose performance across units is not obviously and significantly unbalanced.

2.6 In tiered subjects, where the same grade boundary may feature in two tiers, separate sets of student work for the boundary should be provided from each tier.

In addition for AS/A level specifications:

2.7 Where awarding organisations have to supply student work for an A level specification, two samples are required: one for the AS and one for the A2 units.

2.8 For AS, the work of 15 students whose mark for the AS is at or close to the UMS boundary for an AS grade A (240) or grade E (120) should be supplied. Students

selected should be those whose performance across the three AS units is not obviously or significantly unbalanced. Students should have taken at least two of the three AS units in the June examination series.

2.9 For A level, the sample comprises the A2 work of 15 students who have gained c240 UMS marks at A or c120 UMS marks at E on their A2 units. Students selected should be those whose performance across the three A2 units is not obviously or significantly unbalanced. Students selected will ideally have also gained an overall A level mark which is at or close to the UMS boundary for an overall A level grade A (480) or grade E (240). Students should have taken at least two of the three A2 units in the June examination series.

2.10 The set of AS and A2 units provided should also be a valid combination for A level.

2.11 Where coursework forms a compulsory sub-component within a unit, that coursework should also be collected. Where a unit has optional sub-components, the highest entry option should be supplied. The students chosen for the sample should, as far as possible, have a performance across the components of the unit which is not obviously unbalanced.

Appendix B: Availability of specification materials for the purposes of this review

Materials	AQA	OCR
	Specification	✓
Question paper	✓	✓
Mark scheme	✓	✓
Chief examiner's report	✓	✓
Mark distribution	✓	✓
Grade boundaries	✓	✗
Assessment grids	✓	✗

✓ Material was available and was used in the review

✗ Material was not available and was not used in the review

Appendix C: Student achievement by grade

Cumulative percentage of GCE A level critical thinking grades achieved 2010

Awarding Organisation and level	A*	A	B	C	D	E	U	Total student entries
AQA AS	0.00%	7.20%	20.97%	43.84%	65.32%	84.56%	100.00%	1736
AQA A2	0.79%	5.56%	15.08%	30.16%	57.14%	86.51%	100.00%	126
OCR AS	0.00%	10.84%	29.35%	51.81%	70.74%	84.76%	100.00%	17083
OCR A2	4.17%	13.36%	35.56%	64.33%	86.82%	96.40%	100.00%	1752

Appendix D: Number of data pairs statistically analysed in the script review

Number of data pairs analysed		Number of blank lines	Number of missing/null observations	Number of valid responses used	
Grade	A AS	1232	0	0	1232
	A A2	1232	0	0	1232
	E AS	1232	0	0	1232
	E A2	1120	0	0	1120

Appendix E: Measure, standard error (SE) and infit values of the ranked scripts

The “measure” value represents quality of student performance as judged by the reviewers. It is an estimate of where each script would be ranked if all the scripts were put in order from highest to lowest in terms of performance in a single list. Positive values represent the scripts in the top half of all those reviewed.

The SE is the standard error of the estimated measure value. This is likely to be an underestimate as the analysis changed the rankings (as completed by reviewers on the data-entry sheet for each session) into paired comparisons. The table below exemplifies this. There are four ranking positions. Each rank will be compared against every other position and not just in the order in which they appear.

Reviewer: number 1		Paired comparisons made					
Ranking Position	Script Number	65,23	23,65				
1	65	65,48	48,65	23,48	48,23		
2	23	65,52	52,65	23,52	52,23	48,52	52,48
3	48						
4	52						

Each of the ranked scripts will be paired with each of the other ranked scripts twice for comparison. So, for example, rank 1 will be compared with rank 2 and rank 2 will be compared with rank 1 (hence the paired comparison).

The Infit Z value provides an indication of fit. The higher values indicate that there is more disagreement about the ranking of scripts. For example, scripts that were sometimes ranked above good scripts but at other times ranked below poor scripts (therefore, not consistently positioned within the rankings).

The separation reliability value (infit mean squared) provided is an estimate of the proportion of variance in the script measures attributable to “true” variance as opposed to “error” variance. This is likely to be overestimated, as the analysis changed the rankings into paired comparisons. The separation value, therefore, is how spread the group of measures of the scripts is. The higher the separation value

the better, as this indicates more confidence in the degree of separation between the scripts (that is to say that there is more certainty in the discrimination between them, as observed by the reviewers during the ranking exercise). So the order of the scripts (in terms of the quality of student performance) is more reliable for the sample of scripts reviewed.

Note that the infit mean squared columns' information will always be a positive number (as it has been squared).

The scripts are listed by student performance, with the highest first.

Critical thinking: GCE grade A at AS				
Measure	SE	Awarding organisation	Infit mean squared	Infit Zstd
1.67	0.34	AQA	1.06	0.3
1.52	0.35	AQA	0.89	-0.4
0.96	0.41	AQA	0.97	-0.1
0.96	0.41	AQA	0.87	-0.9
0.95	0.3	AQA	1.05	0.3
0.92	0.28	AQA	0.99	0
0.83	0.45	AQA	1	0
0.83	0.45	AQA	0.98	0
0.81	0.41	AQA	1.02	0.1
0.79	0.27	OCR	1.09	0.9
0.54	0.4	OCR	1.04	0.3
0.4	0.4	AQA	0.88	-0.9
0.3	0.38	AQA	0.88	-0.7
-0.06	0.27	AQA	0.98	-0.2
-0.17	0.45	AQA	1.23	1

Critical thinking: GCE grade A at A2				
Measure	SE	Awarding organisation	Infit mean squared	Infit Zstd
2.67	0.42	AQA	1.03	0.2
2.38	0.56	AQA	1.08	0.3
1.68	0.42	AQA	0.82	-0.7
1.41	0.35	AQA	0.88	-0.6
1.35	0.37	AQA	1.05	0.3
1.32	0.34	OCR	1.12	0.6
1.03	0.38	OCR	1.08	0.5
0.9	0.31	AQA	0.75	-1.8
0.9	0.53	OCR	0.8	-0.5
0.6	0.44	OCR	1.11	0.6
0.3	0.38	AQA	0.94	-0.3
0.24	0.41	AQA	0.99	0
0.14	0.42	AQA	1.14	0.8
0.06	0.39	OCR	1.13	0.8
-0.01	0.42	OCR	1.01	0.1

Review of Standards in GCE A level Critical Thinking: 2010

-0.17	0.45	OCR	1	0
-0.18	0.37	AQA	0.9	-0.7
-0.25	0.28	OCR	1.03	0.3
-0.35	0.27	OCR	1.07	0.6
-0.49	0.4	OCR	0.95	-0.3
-0.49	0.4	OCR	1.01	0.1
-0.67	0.46	OCR	1.15	0.7
-0.74	0.31	OCR	1.04	0.3
-0.78	0.38	OCR	1.08	0.5
-0.8	0.3	OCR	0.98	-0.1
-0.92	0.42	AQA	1	0
-1.07	0.43	OCR	0.98	0
-1.31	0.42	OCR	0.94	-0.2
-1.41	0.5	OCR	0.82	-0.6
-1.62	0.53	OCR	1.02	0.1

-0.01	0.42	OCR	1.25	1.5
-0.04	0.41	OCR	1.36	1.5
-0.2	0.4	OCR	1.11	0.6
-0.29	0.3	OCR	0.99	0
-0.57	0.44	AQA	0.91	-0.3
-0.7	0.34	AQA	1.04	0.3
-0.73	0.44	AQA	0.97	0
-0.9	0.43	OCR	0.81	-0.9
-0.92	0.44	AQA	0.9	-0.4
-1.38	0.47	OCR	0.89	-0.4
-1.38	0.47	AQA	1.15	0.6
-1.45	0.35	OCR	0.99	0
-1.81	0.39	OCR	1.14	0.6
-2.14	0.51	AQA	0.86	-0.4
-2.44	0.41	OCR	0.96	0

Critical thinking: GCE grade E at AS				
Measure	SE	Awarding organisation	Infit mean squared	Infit Zstd
1.34	0.43	AQA	0.95	-0.1
1.24	0.46	AQA	0.86	-0.6
1.07	0.45	AQA	0.95	-0.2
0.74	0.42	AQA	1.07	0.4
0.71	0.31	AQA	0.97	-0.1
0.7	0.3	AQA	1	0
0.6	0.42	OCR	0.99	0
0.57	0.37	OCR	1.15	1.1
0.55	0.39	AQA	1.06	0.3
0.52	0.29	AQA	1.02	0.1
0.32	0.4	AQA	1.02	0.1
0.3	0.42	AQA	0.89	-0.6
0.21	0.27	OCR	1.06	0.7
0.18	0.37	AQA	0.95	-0.2
0.15	0.26	AQA	0.99	0
0.05	0.39	OCR	0.99	0
-0.08	0.4	AQA	1.05	0.4
-0.1	0.36	AQA	1.03	0.2
-0.16	0.27	AQA	1.13	1.4
-0.35	0.27	OCR	1	0

Critical thinking: GCE grade E at A2				
Measure	SE	Awarding organisation	Infit mean squared	Infit Zstd
2.99	1.05	OCR	0.9	0.1
2.87	0.57	AQA	1.06	0.2
2.79	0.55	AQA	1.08	0.3
2.23	0.4	AQA	0.94	-0.2
1.82	0.48	OCR	0.81	-0.8
1.45	0.47	AQA	1.05	0.3
1.16	0.54	OCR	0.91	-0.4
1.16	0.54	AQA	0.91	-0.4
1.16	0.54	AQA	0.91	-0.4
1.13	0.62	AQA	1.31	0.7
0.8	0.5	AQA	0.92	-0.2
0.72	0.48	OCR	0.75	-0.9
0.62	0.37	OCR	1.14	0.7
0.61	0.49	OCR	1.07	0.3
0.23	0.36	AQA	1.11	0.5
0.23	0.48	OCR	1.33	1.3
0.23	0.48	AQA	0.85	-0.6
-0.01	0.49	AQA	1.01	0.1
-0.76	0.58	OCR	0.66	-0.8
-0.89	0.33	OCR	1.01	0

Review of Standards in GCE A level Critical Thinking: 2010

-0.5	0.41	OCR	1.03	0.2
-0.79	0.45	OCR	0.88	-0.4
-0.85	0.37	OCR	1.03	0.2
-0.97	0.38	OCR	1.05	0.4
-0.97	0.38	OCR	0.8	-1.3
-0.98	0.32	OCR	0.87	-0.9
-1.06	0.32	OCR	0.92	-0.4
-1.09	0.31	OCR	1.01	0.1
-1.34	0.3	OCR	1.13	0.9

-1.09	0.46	OCR	0.9	-0.5
-1.16	0.34	AQA	1.02	0.2
-1.19	0.36	OCR	1.28	1.3
-1.65	0.58	AQA	1.13	0.4
-1.85	0.37	AQA	0.93	-0.2
-2.13	0.44	OCR	1.01	0.1
-2.32	0.61	AQA	1	0.1
-2.36	0.44	OCR	0.94	-0.1
-3.23	0.64	OCR	1.06	0.2
-3.54	0.69	OCR	0.91	-0.1

Appendix F: Review team

Review team		Organisation
Lead reviewer	Ruth Matthews	Ofqual reviewer
Specification reviewers	John Chapman	Ofqual reviewer
	Jo Lally	Ofqual reviewer
	James Holiday-Scott	Ofqual reviewer
Script reviewers	John Butterworth	AQA
	Jacque Thwaites	OCR

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

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

© Crown copyright 2012

You may re-use this publication (not including logos) free of charge in any format or medium, under the terms of the [Open Government Licence](#). To view this licence, visit [The National Archives](#); or write to the Information Policy Team, The National Archives, Kew, Richmond, Surrey, TW9 4DU; or email: psi@nationalarchives.gsi.gov.uk

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

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

Office of Qualifications and Examinations Regulation	
Spring Place	2nd Floor
Coventry Business Park	Glendinning House
Herald Avenue	6 Murray Street
Coventry CV5 6UB	Belfast BT1 6DN

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346