



Qualifications and
Curriculum Authority

GCSE (short course) citizenship studies comparability study

March 2007

QCA/07/3089

Contents

Executive summary	4
Background to study	4
Methodology	4
Findings	4
Action plan meeting	5
Introduction	6
Purpose of comparability studies	6
Objectives	6
Qualifications	7
Methodology	8
Summary.....	8
Materials	8
Personnel.....	9
Analytical tools	9
Findings	14
Syllabus review	14
Overall summary.....	14
Content	14
Assessment	15
Overall demand of qualification.....	15
Script review	16
Action plan meeting	19
Purpose.....	19
Presentation of findings.....	19
Implications of findings.....	19
Agreed action points	19
Actions taken	20

Actions taken by awarding bodies in 2006 awarding	20
Conclusion	21
Demand of syllabuses	21
Performance at grade boundaries	21
Awarding in 2006	22
Appendix 1: Reviewers	24
Appendix 2: Awarding body representatives at the Action Plan Meeting	25

Executive summary

Background to study

Following the introduction of citizenship into the national curriculum with effect from August 2002 and the first award of the GCSE (short course) in the summer of 2003, QCA commissioned a review of the short course in the autumn of 2004 (after two awarding cycles) to evaluate its impact on the provision for citizenship and on teaching and learning. This review found sufficient differences between the schemes of assessment, the nature of the examinations and their awards (in terms of raw grade distributions) to recommend that further investigation into the assessment, marking and grading of the short course across the three awarding bodies offering the qualification (AQA, Edexcel and OCR) should be carried out. This comparability study is the result of that recommendation.

Methodology

Six reviewers – three independent consultants appointed by QCA and one representative nominated by each awarding body – carried out a review of the qualifications in two stages. The first stage was a syllabus review, in which the reviewers were provided with syllabus materials from the three awarding bodies and asked to analyse them and make judgements on their level of demand. The second stage was a script review, in which the reviewers assessed candidate work (written papers and coursework) submitted by the awarding bodies and judged the degree of comparability of performance at the key grade boundaries. The script review concluded with a plenary session to draw together findings and issues.

Findings

The outcomes of the syllabus review were as follows.

- The demand of all three qualifications was judged to be 'about right' for the nature, range and number of topics and for the coverage of objectives and the openness of questions.
- AQA was judged to be more demanding for the nature of tasks and the accessibility and targeting of questions.
- Edexcel was judged to be more demanding on the time available per question.¹

¹ Edexcel increased the time allowance for the 2006 examination.

- OCR was judged to be more demanding for the requirements of coursework.

Taken overall, AQA was judged to be the most demanding qualification, with Edexcel and OCR roughly equal in demand.

The outcomes of the script review of written papers were as follows.

- At grades A, C and F, candidates from AQA were judged to be slightly better than candidates from Edexcel and OCR.
- At grades A and C, candidates from OCR were judged to be slightly better than those from Edexcel.
- At grade F, candidates from Edexcel were judged to be slightly better than those from OCR.

A lack of consistency in coursework at all grade boundaries made it very hard to draw confident conclusions about standards, but with that caveat:

- at grade A, work from AQA was judged to be slightly better than that from OCR, with Edexcel in the middle
- at grades C and F, work from AQA was judged to be slightly better than that from Edexcel and OCR, which were broadly in line.

Action plan meeting

Following the review an action plan meeting was held, with the purpose of presenting the findings of the study to representatives of the awarding bodies and agreeing the implications of the findings and any action points arising from them.

At the meeting it was agreed that the 2006 entry was significantly up on 2005, and that a large change, together with a lack of certainty about the nature of the additional candidates, would make the process of grading the summer examinations difficult.

However, it was agreed that the awarding bodies would use the outcomes of the review to inform the awarding process.

Introduction

Purpose of comparability studies

One of the main purposes of QCA's monitoring activities is to ensure that standards of achievement required for a GCSE or A level award meet the regulatory requirements for quality, rigour, fairness and consistency within and across qualifications, across awarding bodies and over time. Comparability studies are one of the principal means by which we fulfil this regulatory function.

The process of a comparability study can be divided into two main sections: syllabus review, which looks at the level of examination demand, and script review, which looks at standards of candidate performance at key grade boundaries. For both sections we draw upon the expertise of subject specialists (reviewers) who have a sound understanding and experience of the assessment system.

For each comparability study the following questions are asked.

- Are the syllabuses equally demanding across all the awarding bodies, and if not, why?
- Was the level of performance required at a given grade boundary equivalent across all the awarding bodies, and if not, how did it differ?

It is important to ensure that:

- all aspects of the examination are covered
- subject specific issues are teased out
- curriculum issues are taken into account.

The final stage of a comparability study is a meeting with suitable awarding body staff to agree an action plan for implementing any changes required.

Objectives

As well as answering the questions asked by every comparability study (see above), the particular objective of this study was to further investigate whether the differences found by the autumn 2004 review had been addressed in the 2005 examinations.

Qualifications

The following qualifications were reviewed:

- AQA citizenship studies (short course) (specification 3107)
- Edexcel citizenship studies (short course) (specification 3280)
- OCR citizenship studies (short course) (specification 1048).

Methodology

Summary

Following an initial briefing meeting, reviewers were sent syllabus materials from the three awarding bodies and asked to analyse the question papers and mark schemes and record their judgements on template forms.

A script review meeting was then held at which the reviewers assessed candidate work submitted by the awarding bodies. This meeting concluded with a plenary session to draw together findings and issues and to produce suggested performance descriptions.

Materials

The following materials were requested from the awarding bodies and provided to the reviewers for the syllabus review:

- specifications
- question papers
- mark schemes
- Chief Examiner's reports
- specification grids for the question papers.

The following materials were requested from the awarding bodies and provided to the reviewers for the script review:

- ten scripts evenly spread from the A* boundary to the B boundary
- ten scripts evenly spread from the B boundary to the D boundary
- ten scripts evenly spread from the E boundary to the G boundary
- ten examples of coursework at the grade boundaries A, C and F.

Personnel

The review used six consultants for both elements. Three were independent consultants appointed by QCA. These consultants had a good general knowledge of assessment together with sufficient subject background to be able to evaluate the materials. The other three comprised one person involved with the qualification at a senior level nominated by each of the three awarding bodies offering the examination.

Analytical tools

Syllabus review

This stage of the work was intended to help reviewers to identify:

- the suitability of question papers for the target candidates and the specification
- how mark schemes contributed to accurate and consistent marking
- the degree of comparability of demand between the awarding bodies
- examples of good practice that should be shared between all awarding bodies
- issues that should be investigated at the script review.

The following analytical tools were used at this stage.

- A form: factual analysis of syllabuses, question papers and mark schemes – one per awarding body; completed by the QCA Project Coordinator with comments added by reviewers.
- Question paper and mark scheme analysis 1st stage (exam paper), incorporating 'CRAS' analysis – one per awarding body; completed by reviewers.
- Question paper and mark scheme analysis 1st stage (coursework/portfolio) – one for all three awarding bodies; completed by reviewers.
- Question paper and mark scheme analysis 2nd stage (summary of judgements) – one for all three awarding bodies; completed by reviewers.

A form

This form gave reviewers a detailed breakdown of each qualification, covering:

- general information (overall page length of syllabus, general philosophy, aims, specification of prior knowledge, additional syllabus support material, anticipated teaching time, access to resources)
- assessment objectives (weighting, how assessed, comments)
- scheme of assessment (component, weighting, time allowance, type of assessment, comments)
- options within assessment (summary of pattern of option availability)
- mark schemes (nature and extent)
- summary (key points related to comparability issues).

Question paper and mark scheme analysis 1st stage (exam paper)

For each specification, this form asked reviewers to comment and make judgements on:

- question papers (overall quality, accessibility, level of demand, fairness to candidates)
- mark schemes (overall quality, discrimination, consistency, accuracy and reliability)
- specification content/coverage (coverage of content and assessment objectives, match of assessment objective weightings and mark allocations)
- differentiation (suitability of question paper and mark scheme for the full range of candidates, range of question types)
- comparability of requirements between the examination paper and the coursework task
- good practice worthy of encouragement and dissemination.

The form also asked reviewers to carry out an individual question analysis, an assessment objective analysis and a CRAS analysis (see below), and to identify specific issues to consider and investigate at the script review.

Question paper and mark scheme analysis 1st stage (coursework/portfolio)

For each specification, this form asked reviewers to comment and make judgements on:

- specification requirements and content (structure and format of the tasks; type and amount of evidence required; particular equipment, resources or demands on centres; penalties, restrictions or word limits; appropriateness of tasks)
- task setting (suitability, accessibility and range of tasks for different candidates; quality of support, guidance and advice available to centres; amount of flexibility or similarity between tasks for different candidates/centres)
- assessment objectives (coverage and balance between relevant assessment objectives for tasks and sub-tasks; range of opportunities for candidates to demonstrate required skills, knowledge and understanding)
- assessment criteria (format/structure of assessment criteria; level of detail and ease of use of criteria; quality of match between specification content, assessment objectives and assessment criteria for tasks and sub-tasks)
- overall quality, validity and reliability of portfolio/coursework tasks.

Question paper and mark scheme analysis 2nd stage (summary of judgements)

This form asked reviewers to rate each awarding body on a scale of 1 to 5, 1 being very undemanding, 3 being 'about right' and 5 being very demanding, in the following areas:

- content: nature of topics, range of topics, number of topics
- assessment: nature of tasks, time available per question, accessibility of questions, coverage of objectives, openness of questions, targeting of questions, overall demand of question papers, requirements of coursework.

It was made clear to reviewers that the purpose of these ratings was not to allow for quantitative analysis, but to reveal patterns of judgements intended to help reviewers to identify significant issues and highlight where further qualitative comments and judgements were required. The form also asked reviewers to provide a summative rank for the awarding bodies on a scale of 1 to 3, 1 being the most demanding and 3 the least demanding, and to list five key points about the analysis they had undertaken.

CRAS analysis

Although the analysis of question papers and mark schemes was essentially qualitative, the CRAS analysis required reviewers to make judgements about each question using numerical values. As before, these judgements were intended to help reviewers to identify significant issues and highlight where further qualitative comments were required.

Reviewers were asked to rate each question, and the question paper as a whole, for the following criteria.

- **Complexity** – the complexity of each constituent operation or idea and the links between them.
- **Resources** – the use of data and information.
- **Abstractness** – the extent to which the candidate is required to deal with ideas rather than concrete objects/phenomena.
- **Strategy** – the extent to which the candidate is required to devise (or select) and maintain a strategy for tackling and answering the question.

Reviewers were asked to use a 4-point scale, where 1 represented a low requirement and 4 a high requirement.

Script review

This stage of the work was intended to help reviewers to identify:

- the degree of comparability between candidates' work at grades A, C and F from the three awarding bodies
- whether examinations were fair and effective for all candidates
- any effects from issues identified in the question paper and mark scheme analysis.

The following analytical tools were used at this stage:

- Form 1: record of judgements from Thurstone pairs exercise
- coursework ranking form.

Review of performance in the written papers

The Thurstone pairs methodology was used to investigate comparability in standards of performance across the awarding bodies' external examinations. The process requires experts to compare one script with another and judge which is the better; the judgements are then analysed to produce statistics showing the relative position of each script used in the exercise. In this case, reviewers were given two packs of candidate work from different awarding bodies, each pack containing ten scripts evenly spread across a range of marks around a given grade boundary. They were then asked to take one script from each pack and decide which of the pair demonstrated a higher level of performance. No ties were allowed. Reviewers worked through the packs until all the scripts had been used in a comparison at least once, with no script being used more than twice.

The reviewers' judgements were then analysed using a Rasch model. The outcome of this analysis enabled the relative performances of the candidates at the respective grade boundaries from each awarding body to be compared.

Review of performance in coursework

Reviewers were given packs of work, each containing samples of coursework at one of the three key grade boundaries from all three awarding bodies. The number of samples from each awarding body varied and was in all cases limited. Reviewers were asked to rank the samples in terms of the overall level of performance they showed, from best to worst. No ties were allowed.

The various ranks were then analysed to determine how far the various samples represented significantly differing standards of performance.

Findings

Syllabus review

Using the analytical tools described in 'Methodology', reviewers commented and made judgements on all aspects of the question papers and mark schemes, and summarised their comments and judgements under the following headings:

- overall summary
- content
- assessment
- overall demand of qualification.

Overall summary

There was consensus among the reviewers on the following points.

- There were more differences between the awarding bodies in the demand of the coursework tasks than in the demand of the question papers.
- The three boards interpreted AO2² differently.
- The difference between Edexcel's time allocation and that of the other two awarding bodies had an impact on the perceived demand of the paper.

Content

Most of the reviewers judged the demand of all three qualifications to be 'about right' for the nature, range and number of topics.

² Assessment objective 2 (AO2) requires candidates to obtain, explain and interpret different kinds of information, including from the media, in order to discuss, form and express an opinion formally, and in writing, and demonstrate their ability to analyse and present evidence on a variety of issues, problems and events.

Assessment

Most of the reviewers judged the demand of all three qualifications to be 'about right' for the coverage of objectives and the openness of questions. Where reviewers consistently judged demand to be other than 'about right', there was consensus on the following points.

- Nature of tasks: AQA was judged to be more demanding because of the type of question used (it was the only paper not to use multiple-choice questions).
- Time available per question: Edexcel was judged to be more demanding because there was less time available for the paper (1h 15min as opposed to 1h 30min for AQA and OCR).³
- Accessibility of questions: AQA was judged to be more demanding because the questions were found to give candidates less support and guidance.
- Targeting of questions: AQA was judged to be more demanding because the wider range of question types used by Edexcel and OCR was found to make them more suitable for the full range of candidates.
- Overall demand of question papers: AQA was judged to be more demanding because of the points made above.
- Requirements of coursework: OCR was judged to be more demanding because task B was found to be more abstract, with limited emphasis on active participation.

Overall demand of qualification

Taking all their judgements into consideration, all the reviewers judged AQA to be the most demanding qualification of the three, with Edexcel and OCR roughly equivalent in demand.

³ Edexcel increased the time allowance to 1h 30min for the 2006 examination.

Script review

Performance at grade boundaries

Written paper

Before reporting the outcomes of the analysis of the judgements made, it is important to note the following factors. There were some significant differences in the level of demand in the question papers. There is substantial evidence that this makes it particularly difficult to make comparative judgements. In addition, the different approaches to the overall structure of the schemes of assessment and, in particular, to the coursework also impact on the extent to which reviewers can consistently judge the merits of work seen.

There is one important additional factor: the Edexcel paper was judged to be very demanding in terms of the time available. This had been recognised by Edexcel and, as already noted, the 2006 paper allowed candidates an extra 15 minutes. However, it is likely that candidates of a particular level of ability will have performed less well on the 2005 paper than might have been expected.

At all three grades, candidates from AQA were judged to be slightly better than candidates gaining the same grade from the other awarding bodies. Candidates from OCR at grades A and C were also judged slightly better than those from Edexcel. At grade F the position between OCR and Edexcel was reversed, with the Edexcel candidates being judged slightly better than those from OCR.

It was noted that the level of correlation between the judgements made and the marks gained was usually quite strong. However, for Edexcel at grade F, there was very little correlation, possibly suggesting that, at this level in particular, the effect of the time available for the question paper was very strong.

Coursework

Perhaps the most obvious point to emerge about the coursework at all grade boundaries was the lack of consistency found in candidate work within awarding bodies. This was particularly the case with OCR, where reviewers often commented on the difference in standard shown across the two pieces of work produced. The level of inconsistency may also derive from problems with moderation for all of the awarding bodies. It also makes it very hard to draw confident conclusions about standards between the awarding bodies.

At grade A, insofar as it was possible to judge, work from AQA was slightly better than that from OCR, with Edexcel in the middle. At grades C and F, work from AQA was again judged slightly better than that from the other two awarding bodies, which were broadly in line.

Performance descriptions

At the end of the script review, reviewers were asked to suggest amendments to the published performance descriptions in the light of the candidate work they had observed. Their amendments were then incorporated into the performance descriptions, as follows (additions are underlined, deletions are struck through).⁴

Grade A

Candidates demonstrate in-depth knowledge and understanding about ~~rights and responsibilities, communities and identities, democracy and government~~ citizenship topics and issues and the impact these have on societies, culture and the global community. They do this by drawing on specific examples and explaining the meaning and importance of citizenship issues and concepts in relation to their own and other people's lives and their communities. They discuss, interpret and evaluate a variety of different ~~responses~~ sources of information demonstrating an appreciation of other points of view. They recognise the complexity of issues studied, weighing up opinions and make judgements supported by a range of evidence and well-developed arguments. They critically evaluate their participation in school and/or community citizenship activities ~~providing evidence of~~ and their ability to work with others and place this within a wider context. They show they can make a range of contributions to group decision-making and critically assess the effects of their participation. ~~They have negotiated their roles and responsibilities within project work groups, and facilitated the participation of others.~~

Grade C

Candidates demonstrate sound knowledge and understanding about rights about ~~rights and responsibilities, communities and identities, democracy and government~~ citizenship topics and issues and the impact these have on societies, culture and the global community. They do this by describing and explaining the importance and influence of citizenship issues in relation to their own and others people's lives and their communities. They discuss, interpret and evaluate different responses to issues studied by presenting relevant evidence to support

⁴ These amendments were passed on to the awarding bodies to be used to inform the awarding process; they were not formally adopted into the published performance descriptions.

arguments. They demonstrate an appreciation of different points of view and use arguments to make reasoned judgements. They evaluate their participation in school and/or community citizenship activities providing evidence of their ability to work with others, contributing effectively to group decision making and project outcomes.

Grade F

Candidates demonstrate elementary knowledge and basic understanding about ~~rights and responsibilities, communities and identities, democracy and government~~ citizenship topics and issues and how these impact on societies, culture and the global community. They do this by using their knowledge to make simple connections between ~~their understanding of~~ citizenship issues in relation to their own and other people's lives, and their communities. They may take part in discussions and can express opinions ~~with reasons~~. They provide evidence that they have taken part with others in school and/or community citizenship activities ~~and reflected on their participation~~ by describing their role.

Exemplar scripts

Reviewers were also asked to select scripts which represented a typical performance at each grade; these were retained as exemplar scripts for future reference.

Action plan meeting

Purpose

The purpose of the action plan meeting was to present the findings of the study to representatives of the three awarding bodies offering the qualification and to agree the implications of the findings and any action points arising from them.

Presentation of findings

Script review

It was agreed that the 2006 entry was significantly up on 2005, and that a large change, together with a lack of certainty about the nature of the additional candidates, would make the process of grading the summer examinations difficult.

For Edexcel, the situation was further complicated by the change in examination time and uncertainty about exactly how this would affect candidates at different levels of attainment.

Coursework

Awarding body representatives thought that centres were devoting insufficient resources to coursework, and against that background it was difficult for them to encourage improvement in the quality of coursework. It was also felt that coursework in citizenship represented the outcome of candidates' active role in the community and as such should be seen as a core strength of the qualification.

Implications of findings

Awarding body representatives were asked to be aware of the differences between their respective qualifications. It was noted that the involvement of senior examining personnel in all stages of the process, including amending the performance descriptions and identifying exemplar scripts, would be of assistance in this.

Agreed action points

It was agreed that the awarding bodies would use the exemplar scripts selected at the end of the script review, and the suggested amendments to the performance descriptions when they became available, to inform the awarding process.

Actions taken

Actions taken by awarding bodies in 2006 awarding

Aware of low outcomes in 2005 and the findings of this comparability study, AQA was able to award a higher proportion of grades C and F . AQA also noted an improvement in coursework outcomes.

The additional allowance of 15 minutes made by Edexcel for the 2006 examination, to bring it into line with the other two awarding bodies, resulted in an increase in the mean mark of 11 raw marks; Edexcel raised the raw mark boundaries to reflect this.

OCR took the inter-awarding body data into account in reviewing and amending the key grade boundaries.

Conclusion

At the start of the study the following questions were asked.

- Are the syllabuses equally demanding across all the awarding bodies, and if not, why?
- Was the level of performance required at a given grade boundary equivalent across all the awarding bodies, and if not, how did it differ?

Demand of syllabuses

The outcomes of the syllabus review were as follows.

- The demand of all three qualifications was judged to be 'about right' for the nature, range and number of topics, and for the coverage of objectives and the openness of questions.
- AQA was judged to be more demanding for the nature of tasks and the accessibility and targeting of questions.
- Edexcel was judged to be more demanding for the time available per question.
- OCR was judged to be more demanding for the requirements of coursework.

Taken overall, AQA was judged to be the most demanding qualification, with Edexcel and OCR roughly equivalent in demand.

Performance at grade boundaries

Written papers

Allowing for the differences in the level of demand of the question papers, and the difference in the time available for the Edexcel paper, the outcomes of the script review of written papers were as follows.

- At grades A, C and F, candidates from AQA were judged to be slightly better than candidates from Edexcel and OCR.
- At grades A and C, candidates from OCR were judged to be slightly better than those from Edexcel.

- At grade F, candidates from Edexcel were judged to be slightly better than those from OCR.

Coursework

Reviewers found a lack of consistency in coursework from all three awarding bodies at all grade boundaries. This made it very hard to draw confident conclusions about standards, but with that caveat:

- at grade A, work from AQA was judged to be slightly better than that from OCR, with Edexcel in the middle
- at grades C and F, work from AQA was judged to be slightly better than that from Edexcel and OCR, which were broadly in line.

Awarding in 2006

At the action plan meeting, it was agreed that the awarding bodies would use the suggested amendments to the performance descriptions and the exemplar scripts selected at the end of the script review to inform the awarding process.

Results in 2005–6

		Percentage of candidates at each grade (cumulative)		
Awarding body	Year	A	C	F
AQA	2005	9.8	39.3	82.7
	2006 ⁵	9.4	40.5	85.5
Edexcel	2005	16.8	59.5	90.2
	2006 ⁶	14.6	56.3	90.0
OCR	2005	10.0	61.1	96.2
	2006 ⁷	10.6	56.7	93.9

⁵ Aware of low outcomes in 2005 and the findings of this study, AQA was able to award a higher proportion of grades C and F without needing to lower the boundary marks.

⁶ The additional allowance of 15 minutes made by Edexcel for the 2006 examination to bring it into line with the other two awarding bodies resulted in an increase in the mean mark of 11 raw marks; Edexcel raised the raw mark boundaries to reflect this.

⁷ OCR took the inter-awarding body data into account in reviewing and amending the key grade boundaries.

Appendix 1: Reviewers

Cate Amsdorf

Peter Brett (AQA nominee)

Anthony Glachan

Trevor Green (Edexcel nominee)

Steven Johnson (OCR nominee)

Carole Waugh

Appendix 2: Awarding body representatives at the Action Plan Meeting

AQA Christina Alley
 Neil Stringer

Edexcel Trevor Green
 Jim Sinclair

OCR Elizabeth Gray
 Stuart Lunn