



Qualifications and
Curriculum Authority

GCSE French comparability study

March 2007

QCA/07/3098

Contents

Executive summary.....	3
Background to study	3
Methodology	3
Findings	4
Action plan meeting	6
Results 2004–06.....	6
Introduction.....	7
Purpose of comparability studies	7
Background to study	7
Objectives.....	8
Qualifications	9
Methodology.....	10
Summary	10
Materials	10
Personnel	10
Analytical tools.....	11
Findings.....	16
2005 study – summary of findings	16
2006 study	16
Action plan meeting.....	24
Purpose	24
Presentation of findings	24
Implications of findings	24
Agreed action points.....	24
Actions taken.....	25
Actions taken by awarding bodies in 2006 awarding	25
Results in 2006.....	25
Conclusion.....	27
Demand of syllabuses	27
Performance at grade boundaries	27
Speaking.....	27
Awarding in 2006.....	28
Appendix 1: Reviewers.....	29
Appendix 2: Awarding body representatives at the action plan meeting.....	30

Executive summary

Background to study

Qualifications in modern foreign languages have had a very high profile for some years. This is partly because uptake is deemed very important for the country's welfare. In this context, it is therefore particularly important that inter-awarding body standards within any qualification are consistent. In 2004, concerns were expressed that this was not in fact the case for GCSE French. Most of the concerns were based on the fact that raw grade distributions were very different across the different awarding bodies, with results for AQA, the market leader, markedly poorer than the others. However, results using concurrent attainment as a proxy measure for general ability suggested that the differences were more apparent than real.

Nevertheless, as a result of the concerns, it was decided to mount a comparability study to investigate whether or not standards were comparable across the various syllabuses. However, because the more sophisticated statistical analyses suggested the problem was much smaller than the raw data suggested, the work was limited to the reading and writing papers from the three English awarding bodies only.¹

In 2005 results were closer together, but there remained a good deal of public concern. In addition, raw grade distributions for both the CCEA and WJEC syllabuses were also markedly different from the English average. It was therefore decided to repeat the exercise, this time looking at all four skills from all five mainstream syllabuses.² For the remainder of this report, the focus will be on the 2005 work, drawing on the earlier work as necessary.

Methodology

Fourteen reviewers – four independent consultants appointed by the Qualification and Curriculum Authority (QCA) and two representatives 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 five 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 submitted by the awarding bodies and judged the degree of comparability of performance at the key grade

¹ Reading and writing were considered much the most useful in terms of making judgements about performance.

² In addition, AQA runs a modular syllabus. However, this would create significant problems for those making judgements, while there is good evidence that standards between the two AQA syllabuses are well aligned.

boundaries A, C and F. The review covered all three key grade boundaries for reading, writing, listening, speaking and writing coursework. The script review concluded with a plenary session to draw together findings and issues.

Findings

The outcomes of the syllabus review were as follows.

- Taken overall, AQA and OCR were judged to be the most demanding qualifications and CCEA and WJEC the least demanding.
- The demand of all five qualifications was judged to be 'about right' for the nature and number of topics, grammatical requirements, time available per question and coverage of objectives.

AQA was judged by some reviewers to be more demanding for:

- the nature of tasks, length/range and demand of texts used as stimulus material, range/demand of vocabulary/grammar in the questions, accessibility of questions for higher attaining candidates, targeting of questions, overall demand of question papers.

CCEA was judged by some reviewers to be less demanding for:

- the nature of tasks, length/range and demand of texts used as stimulus material, range/demand of vocabulary/grammar required to answer the questions, accessibility of questions for higher attaining candidates, openness of questions, requirements of speaking, overall demand of question papers.

It was judged to be more demanding for:

- the range/demand of vocabulary/grammar in the questions, accessibility of questions for lower attaining candidates.

Edexcel was judged by some reviewers to be less demanding for:

- the requirements of speaking.

It was judged to be more demanding for:

- the length/range and demand of texts used as stimulus material.

OCR was judged by some reviewers to be more demanding for:

- the nature of tasks, requirements of speaking.

WJEC was judged by some reviewers to be less demanding for:

- the range of topics, nature of tasks, length/range and demand of texts used as stimulus material, range/demand of vocabulary/grammar in the questions, accessibility of questions for higher attaining candidates, accessibility of questions for lower attaining candidates, targeting of questions, requirements of speaking, overall demand of question papers.

It was judged to be more demanding for:

- the openness of questions, requirements of coursework.

The outcomes of the review of candidates' work were as follows.

- In the higher tier listening paper, AQA was slightly severe at both grade boundaries and CCEA slightly lenient. In the foundation tier paper, there was little difference between any of the awarding bodies at either grade boundary.
- In the higher tier reading paper, there was little difference between the awarding bodies at A, but AQA was slightly severe at C. At foundation tier, there was little difference at grade C, but AQA was slightly severe and CCEA slightly lenient at grade F.
- In the higher tier writing paper, there was little difference between the awarding bodies at grade A, but CCEA was slightly severe at grade C. At foundation tier, there was little difference at either grade boundary.

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, WJEC work was judged to be slightly weaker than that from the other awarding bodies³
- at grade C, there was little difference between the awarding bodies
- at grade F, WJEC work was judged to be slightly weaker than that from the other awarding bodies.

There were two factors about the orals which made it very hard to draw confident conclusions about standards: first there was a lack of consistency at all grade boundaries; second the quality of many of the tapes was very poor. With these caveats:

- at grade A, WJEC work was judged to be slightly weaker and AQA slightly better than that from the other awarding bodies

³ It should be noted that all WJEC GCSE French coursework is done under controlled conditions.

- at grade C (higher tier), WJEC work was judged to be slightly weaker and CCEA slightly better than that from the other awarding bodies
- at grade C (foundation tier), CCEA work was judged to be slightly weaker and AQA slightly better than that from the other awarding bodies
- at grade F, there was very little comparability between the awarding bodies, although this was almost certainly more an issue of the quality of the materials considered than of the standards applied.

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 down on 2005, and that a large change, together with lack of certainty about the nature of the candidates who had dropped out, would make the process of grading the summer's examinations difficult.

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

Results 2004–06

Over the period of the exercise, there was a significant reduction in the differences in raw grade distributions between the awarding bodies. There is no reason to believe that the differences that remain are attributable to anything other than differences in the nature of the candidates entered for each examination.

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 principle means by which QCA fulfils 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?

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.

Background to study

As a result of concerns raised about the consistency of standards across awarding bodies in the 2004 GCSE examinations in modern foreign languages, QCA carried out an investigation into the comparability of GCSE French examinations in that year. Partly because statistical analyses suggested that the differences might be less significant than differences in the raw

grade boundaries implied, the study was confined to the reading and writing papers of the three English awarding bodies.⁴

The 2005 work suggested that there were differences in grading standards across the awarding bodies. The findings of the study were discussed with the awarding bodies so that their implications could be fed into awarding meetings in summer 2005, in compliance with QCA's Code of practice, paragraph 6.14.xii.

However, the awards in 2005 posed several problems in their own right. In particular, the subject of French as a whole, and several of the individual specifications, saw a significant decline in entry. This made it particularly difficult to evaluate the nature of the cohort and whether it had changed in character as well as size from 2004. In addition, although several of the strands of the work were able to supply some measure of the size of the perceived differences in standard, none could give a real indication of whether the standard that any awarding body was applying was appropriate, except in the simplest sense of being closest to the average.

The awards in 2005 did see a significant closing of the differences in raw grade distributions from the awarding bodies. It remained difficult, however, to be confident that the differences identified in the study had been fully resolved. It was therefore recommended that there should be further evaluation of standards applied in 2005, this time addressing all the skills assessed and including CCEA and WJEC as well as the three English awarding bodies.

Objectives

As well as answering the questions asked by every comparability study (see above), the particular objective of the 2006 study was to attempt to reach consensus on the level of performance appropriate to each grade reviewed in each skill. A key product of this part of the exercise would be a set of performance descriptions for performance at grades A, C and F by skill, together with examples of work at the agreed standard.

⁴ It was also because reading and writing were considered the most useful in terms of making judgements about performance.

Qualifications

The following qualifications were reviewed:

- AQA GCSE French 3651 specification A (full course)
- CCEA GCSE French 5650
- Edexcel GCSE French 1226
- OCR GCSE French 1925
- WJEC GCSE French 0157 (full course).

Methodology

Summary

Following an initial briefing meeting, reviewers were sent syllabus materials from the five 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. Because of the quantity of work involved in the script review, a further meeting was held 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
- specification grids for the question papers
- tapes and transcripts for the listening paper
- tasks for orals
- mark schemes
- chief examiner's reports.

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

- higher tier – for each of the reading, writing and listening tests:
 - ten scripts evenly spread from the A* boundary to the B boundary
 - ten scripts evenly spread from the B boundary to the D boundary
- foundation tier – for each of the reading, writing and listening tests:
 - ten scripts evenly spread from the notional maximum mark to the D boundary
 - ten scripts evenly spread from the E boundary to the G boundary
- examples of work at the grade boundaries A, C and F for speaking and coursework, up to a maximum of ten candidates.

Personnel

The review used 14 consultants for both elements. Four 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 10 comprised two people involved with the qualification at a senior level nominated by each of the five 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 QCA project coordinator with comments added by reviewers)
- question paper and mark scheme analysis 1st stage (exam paper, for reading, writing and listening) (incorporating 'CRAS' analysis) (one per awarding body, completed by reviewers)
- question paper and mark scheme analysis 1st stage (speaking) (one for all five awarding bodies, completed by reviewers)
- question paper and mark scheme analysis 1st stage (coursework/portfolio) (one for all five awarding bodies, completed by reviewers)
- question paper and mark scheme analysis 2nd stage (summary of judgements) (one for all five 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 reading, writing and listening)

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 target 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 (speaking)

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

For each specification, these forms 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, grammatical requirements (from specification)
- assessment: nature of tasks, time available per question, length/range and demand of texts used as stimulus material on question papers, range/demand of vocabulary/grammar used in the questions, range/demand of vocabulary/grammar required to answer the questions, accessibility of questions for higher attaining candidates, accessibility of questions for lower attaining candidates, coverage of objectives, openness of questions, targeting of questions, overall demand of question papers, requirements of speaking, 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 1 to 5, 1 being the most demanding and 5 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 four-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 (higher tier), C (foundation tier) and F from the five 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 (reading, writing and listening)

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 speaking and coursework

Reviewers were given packs of work, each containing samples of coursework/oral tapes 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.

Follow-up meeting

Because of the quantity of work involved in the script review, a follow-up meeting was necessary to complete this stage of the study. At this meeting reviewers were asked to:

- identify scripts at the right standard
- identify grade, skill and performance descriptions
- set performance in the context of question demand.

Findings

2005 study – summary of findings

Syllabus review

In the 2005 study, reviewers judged the overall demand of all three qualifications⁵ to be comparable; they agreed that overall all three would be demanding of candidates at both foundation and higher tiers. More detailed findings were as follows:

- **stimulus materials:** reviewers judged that the source material for the AQA reading papers used particularly sophisticated and complex language, and that in terms of length of text, the OCR reading papers appeared demanding at both foundation and higher tiers
- **range/demand of vocabulary:** reviewers judged that across all three qualifications the vocabulary requirements were extremely demanding for all candidates; they also judged that AQA and OCR were slightly more demanding because of the use of some less familiar vocabulary
- **range/demand of grammar:** reviewers judged that the grammatical requirements of all three qualifications were too demanding for all candidates; they also judged that the OCR questions were more demanding because some tasks involved lengthy and complex instructions in the target language
- **accessibility of questions:** reviewers judged that tasks across all awarding bodies were demanding for lower attaining candidates; in particular, they judged that papers from all awarding bodies were very demanding for grade F candidates
- **sequencing of questions/tasks:** reviewers judged that AQA and Edexcel papers could be more demanding because the questions/tasks were not sequenced by demand.

2006 study

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

⁵ AQA, Edexcel and OCR, reading and writing papers only.

- overall demand of qualification.

Overall summary

There was consensus among the reviewers on the following points:

- All five specifications showed a remarkable degree of similarity in terms of content, weighting of the four assessment objectives and grammatical requirements.
- Foundation tier papers were generally well targeted at candidates of this ability. At higher tier there was more of a divergence, with some awarding bodies setting demanding reading and (especially) listening material and more demanding question types.
- There were issues of comparability with regard to speaking tests and coursework.
- Some visuals were not clear and could confuse candidates. Awarding bodies should ensure that all stimuli are unambiguous and can be clearly made out before examination papers are printed.

Content and assessment

The demand of all five qualifications was judged to be 'about right' for the nature and number of topics, grammatical requirements, time available per question and coverage of objectives.

AQA was judged by some reviewers to be more demanding for:

- the nature of tasks, length/range and demand of texts used as stimulus material, range/demand of vocabulary/grammar in the questions, accessibility of questions for higher attaining candidates, targeting of questions, overall demand of question papers.

CCEA was judged by some reviewers to be less demanding for:

- the nature of tasks, length/range and demand of texts used as stimulus material, range/demand of vocabulary/grammar required to answer the questions, accessibility of questions for higher attaining candidates, openness of questions, requirements of speaking, overall demand of question papers.

It was judged to be more demanding for:

- the range/demand of vocabulary/grammar in the questions, accessibility of questions for lower attaining candidates.

Edexcel was judged by some reviewers to be less demanding for:

- the requirements of speaking.

It was judged to be more demanding for:

- the length/range and demand of texts used as stimulus material.

OCR was judged by some reviewers to be more demanding for:

- the nature of tasks, requirements of speaking.

WJEC was judged by some reviewers to be less demanding for:

- the range of topics, nature of tasks, length/range and demand of texts used as stimulus material, range/demand of vocabulary/grammar in the questions, accessibility of questions for higher attaining candidates, accessibility of questions for lower attaining candidates, targeting of questions, requirements of speaking, overall demand of question papers.

It was thought to be more demanding for:

- the openness of questions, requirements of coursework.

Overall demand of qualification

Taking all their judgements into consideration, the reviewers judged AQA and OCR to be the most demanding qualifications, and CCEA and WJEC the least demanding.

It is worth noting that these findings are broadly in line with the earlier exercise.

Script review

Performance at grade boundaries

Written papers (reading, writing and listening)

The outcomes of the review of candidates' work were as follows.

- In the higher tier reading paper, there was little difference between the awarding bodies at A, but AQA was slightly severe at C. At foundation tier, there was little difference at grade C, but AQA was slightly severe and CCEA slightly lenient at grade F.
- In the higher tier writing paper, there was little difference between the awarding bodies at grade A, but CCEA was slightly severe at grade C. At foundation tier, there was little difference at either grade boundary.

- In the higher tier listening paper, AQA was slightly severe at both grade boundaries and CCEA slightly lenient. In the foundation tier paper, there was little difference between any of the awarding bodies at either grade boundary.

The findings are broadly in line with the earlier exercise, although the differences have been reduced. This suggests that the English awarding bodies had successfully improved the alignment of their standards as a result of the first investigation.

Speaking

There were two factors about the orals which made it very hard to draw confident conclusions about standards: first, there was a lack of consistency at all grade boundaries; second, the quality of many of the tapes was very poor. With these caveats:

- at grade A, WJEC work was judged to be slightly weaker and AQA slightly better than that from the other awarding bodies
- at grade C (higher tier), WJEC work was judged to be slightly weaker and CCEA slightly better than that from the other awarding bodies
- at grade C (foundation tier), CCEA work was judged to be slightly weaker and AQA slightly better than that from the other awarding bodies
- at grade F, there was very little comparability between the awarding bodies, although this was almost certainly more an issue of the quality of the materials considered than of the standards applied.

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, WJEC work was judged to be slightly weaker than that from the other awarding bodies⁶
- at grade C, there was little difference between the awarding bodies
- at grade F, WJEC work was judged to be slightly weaker than that from the other awarding bodies.

Performance descriptions

At the follow-up meeting reviewers were asked to develop performance descriptions by suggesting amendments to the published grade descriptions in the light of the candidates

⁶ It should be noted that all WJEC GCSE French coursework is done under controlled conditions.

work they had observed.⁷ Their suggestions were grouped in assessment objectives and subsequently edited and adapted to cover all languages. They describe performance on the grade boundary in terms of what candidates should characteristically be able to do at that level, and also describe how candidates typically perform in parts of the paper not targeted at them. This is particularly relevant for grade C at both tiers. For writing, the descriptions do not separate performance in the external examination and coursework, but there is an extra descriptor, which applies only to coursework.

AO1 listening

At grade A, candidates characteristically:

- identify and note almost all of the details required in ‘overlap’ tasks
- respond to more extended passages of complex language with some success
- identify attitudes, opinions and points of view sometimes from a number of speakers
- are able to sometimes draw conclusions and make inferences
- demonstrate more global understanding of longer passages.

At grade C, candidates characteristically:

- on foundation tier papers, understand and identify single items and detail from short spoken texts via straightforward tasks over a range of contexts, mostly successfully
- on ‘overlap’ tasks, identify details including simple opinions from lengthier passages and dialogues which contain a variety of structure and may include more than one tense and opinions; they may also respond at times in the target language, but with a number of errors
- on higher-tier papers, identify a few familiar items of language or vocabulary in more complex passages and less familiar contexts.

At grade F, candidates characteristically:

- identify or match key words or basic familiar vocabulary from short, spoken texts with some success
- identify occasional single items of familiar vocabulary from longer passages and dialogues and may score marks randomly in more complex objective test types
- show little evidence of identifying meaning where what they hear does not directly match what is seen on the page (eg where synonyms rather than identical words are used).

⁷ It is to be expected that there would need to be changes: a grade description identifies the sort of performance a candidate typical of the grade might show; a performance description is intended to describe performance at the bottom of the grade.

AO2 speaking

At grade A, candidates characteristically:

- take part in role-play tasks almost entirely successfully and are usually able to respond to unpredictable questions
- extend their responses naturally and routinely using a variety of vocabulary
- express and justify opinions
- have a certain natural flow and interaction with the examiner
- use a range of structures including verb tenses more accurately than inaccurately
- attempt more complex language, although not always successfully
- speak quite confidently and attempt genuine pronunciation and intonation.

At grade C, candidates characteristically:

- take part in straightforward role-play tasks mostly successfully
- respond to some unfamiliar prompts or questions, but may be more hesitant in their responses or not understand the question
- respond to familiar questions and sometimes go beyond a minimal or basic response
- communicate a range of information about themselves and their own experiences
- use a range of simple structures, including some variety of tense where appropriate, although there will be errors
- communicate with comprehensible pronunciation and intonation, but may be influenced by their first language at times.

At grade F, candidates characteristically:

- have some success in simple role-play tasks, mostly responding in single words and phrases
- communicate some basic information in brief sentences and phrases but depend heavily on the teacher's prompts
- answer some straightforward familiar questions often by substituting words or phrases but show little evidence of ability to go beyond basic responses
- use language which contains frequent errors and is often strongly influenced by their first language
- may have difficulties with pronunciation which can impede communication
- may produce an uneven performance overall which is stronger in some sections, particularly in pre-prepared tasks such as a presentation, and very weak in others.

AO3 reading

At grade A, candidates characteristically:

- identify and note almost all of the details required in 'overlap' tasks
- respond to more extended texts containing complex and less familiar language with some success
- identify attitudes, opinions and points of view and are able to sometimes draw conclusions and make inferences
- show understanding of grammar (eg tenses) as it affects comprehension
- demonstrate good general understanding or ability to understand the gist.

At grade C, candidates characteristically:

- on foundation tier papers, can deal with straightforward tasks, identifying single items and detail from short simple texts over a range of contexts, almost entirely successfully
- on 'overlap' tasks, identify more details from lengthier texts which contain a variety of structure and may include more than one tense and opinions; they may respond at times in the target language, but with a number of errors
- on higher-tier papers, identify a few familiar items of language or vocabulary in longer and more complex texts and less familiar contexts.

At grade F, candidates characteristically:

- identify basic items of vocabulary and familiar short phrases in lists or short texts with some success
- may be able to identify familiar words or simple phrases in longer or denser texts
- show no evidence of using grammatical knowledge or understanding to discern meaning and may score marks randomly in more complex objective test types.

AO4 writing

At grade A, candidates characteristically:

- produce extended pieces of writing sometimes for different purposes
- express and justify opinions
- use a range of straightforward structures and vocabulary, including tenses as required, mostly appropriately and successfully
- attempt to use more complex structures, although they may not always be successful
- produce some accurate sequences of language
- in coursework use a wider range of more complex language with more accuracy.

At grade C, candidates characteristically:

- can convey simple information in short sentences with some accuracy
- produce more extended writing requiring a wider range of context and structures, including different tenses
- communicate a range of information; although there are mistakes, what they write is mostly comprehensible
- express simple opinions, sometimes with reasons
- write at greater length in coursework, with a wider range of language and with greater accuracy appropriate to the conditions required.

At grade F, candidates characteristically:

- communicate in writing using simple phrases and sentences, despite frequent inaccuracies and interference from their first language
- can sometimes substitute words and phrases to vary sentences
- in coursework can produce longer pieces of writing which although often repetitive and reliant on pre-learnt language, is generally comprehensible despite much inaccuracy.

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 five 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

Awarding body representatives pointed out that the awarding process was influenced by awarding bodies wanting to achieve the right overall outcome in setting grade boundaries rather than by 'getting it right' for each skill.

They also pointed out that an awarding body's papers could be demanding but its awarding lenient (and vice versa), thus balancing out the outcomes in terms of candidates' grades.

Implications of findings

Awarding body representatives were asked to be aware of the differences between their respective qualifications.

Agreed action points

It was agreed that the process of grading the summer's examinations would be difficult. The 2006 entry was significantly lower than in 2005, and it was impossible to be certain about the nature of the candidates who had dropped out.

However, it was agreed that the awarding bodies would use information from the script review together with the suggested performance descriptions and the exemplar scripts selected at the follow-up meeting to inform the awarding process.

Actions taken

Actions taken by awarding bodies in 2006 awarding

Following an instruction from CCEA's principal examiner to apply the mark scheme more rigorously to speaking (higher tier) papers, there was a drop in mean performance. The A and C grade boundaries were accordingly dropped by six and five marks respectively, reflecting a slightly weaker candidature.

The Edexcel awarding committee felt that the increased percentages at grade C reflected the enormous amount of energy expended in schools throughout the country to 'convert' grade Ds to grade Cs, often to the detriment of grade A candidates.

Following an unexpected two-mark rise in the grade C boundary for writing coursework in 2005, OCR again raised the boundary, this time by one mark.

Results in 2006

Table 1 shows how results changed for the five awarding bodies over the period of the study. AQA awarded just under 6 per cent more As in 2006 than in 2004, and well over 13 per cent more Cs. For CCEA and WJEC, outcomes stayed very stable at both grades between 2004 and 2006. Edexcel had slightly better results at grade C in 2006 than in 2004, with outcomes at grade A very stable. For OCR, there was a small improvement at both grades A and C over the period. Overall, this meant that the gap between AQA and all the other awarding bodies was substantially lower in 2006 than 2004.

Table 1 Results in 2005–06

		Percentage of candidates at each grade (cumulative)		
Awarding body	Year	A	C	F
AQA (French A)	2004	16.9	49.8	93.2
	2005	20.4	58.4	95.2
	2006	22.7	63.3	96.8
CCEA	2004	40.8	80.8	98.7
	2005	41.7	81.3	98.8
	2006	41.2	80.2	99.5
Edexcel	2004	23.3	61.4	95.7
	2005	23.2	62.9	97.6
	2006	23.9	66.8	97.9
OCR	2004	24.7	70.2	98.0
	2005	25.7	70.1	98.8
	2006	27.6	72.7	98.7
WJEC	2004	30.2	72.6	98.2
	2005	30.2	73.8	98.6
	2006	30.5	73.3	98.6

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?

Demand of syllabuses

The outcomes of the syllabus review were as follows:

- Taken overall, AQA and OCR were judged to be the most demanding qualifications and CCEA and WJEC the least demanding.
- The demand of all five qualifications was judged to be 'about right' for the nature and number of topics, grammatical requirements, time available per question and coverage of objectives. The reviewers judged there to be difference in demand between awarding bodies in all other elements of the assessment tools.

Performance at grade boundaries

Written papers

The outcomes of the script review of written papers were:

- AQA was slightly severe at both grade boundaries of the higher tier listening paper and at grades C and F of the foundation tier reading paper
- CCEA was slightly lenient at both grade boundaries of the higher tier listening paper, slightly severe at grade C of the higher tier writing paper, and slightly lenient at grade F of the foundation tier reading paper
- there was little difference between the awarding bodies at either grade boundary of the foundation tier listening paper, at grade A of the higher tier and grade C of the foundation tier reading papers and at grade A of the higher tier and both grades of the foundation tier writing papers.

Speaking

The lack of consistency at all grade boundaries and the poor quality of many of the tapes made it very hard to draw confident conclusions about standards. With these caveats:

- AQA work was judged to be slightly better at grade A and grade C (foundation tier) than that from the other awarding bodies

- CCEA work was judged to be slightly better at grade C (higher tier) and slightly weaker at grade C (foundation tier) than that from the other awarding bodies
- WJEC work was judged to be slightly weaker at grade A and grade C (higher tier) than that from the other awarding bodies
- there was very little comparability between the awarding bodies at grade F, although this was almost certainly more an issue of the quality of the materials considered than of the standards applied.

Coursework

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

- WJEC work was judged to be slightly weaker than that from the other awarding bodies at grades A and F⁸
- there was little difference between the awarding bodies at grade C.

Awarding in 2006

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

As a result of awarding decisions taken in 2006, the differences in raw grade distributions, which had already narrowed between 2004 and 2005, were further reduced. There is no reason to believe that the remaining differences arise from anything other than differences in the nature of the candidates entered for each examination.

⁸ It should be noted that all WJEC GCSE French coursework is done under controlled conditions.

Appendix 1: Reviewers

Dave Carter (AQA nominee)

Mary Culpan

David Forth (OCR nominee)

Trevor Halliday (CCEA nominee)

Jill Llewellyn Williams (WJEC nominee)

Richard Marsden (Edexcel nominee)

David Martin (Edexcel nominee)

Tom Moore (CCEA nominee)

Andrew Portas

Janet Searle

Geoff Shooter (AQA nominee)

David Thomas (WJEC nominee)

Caroline Woods (OCR nominee)

Gary Woods

Appendix 2: Awarding body representatives at the action plan meeting

AQA Judith Rowland-Jones
 Neil Stringer

CCEA Sean MacNia
 Brendan McDowell

Edexcel David Martin
 Jim Sinclair

OCR Alec Johns
 Christine Peakman

WJEC Sheila Bullough
 Liz Phillips