Investigating Ways to Improve Written GCSE Examinations

Final Report from the Centre for Developing and Evaluating Lifelong Learning, University of Nottingham

Professor Roger Murphy,
Professor David Greatbatch
John Wilmut
Professor Harry Tolley
Carole Mallia
John Hamer

January 2008

Contents

Page

Acknowledgements

The CDELL team especially wishes to acknowledge the central contribution to the project made by the members of the three subject scrutiny panels (whose names are listed on pages 14-15). We are most grateful to them for the time that they gave during August and September 2007 to both individual and team scrutiny work.

We also wish to acknowledge the help given by Dennis Opposs, Angus Alton and Douglas Lee of QCA. The team is also grateful for the help of the awarding bodies in identifying subject experts to conduct the scrutinies and in providing the examination materials for use in the scrutiny process.

Finally, we would like to thank participants in the consultation conference organised by QCA in December 2007 for their insightful comments and feedback on our interim report.

Executive Summary

Description of the study

- This Project was conducted for QCA by the Centre for Developing and Evaluating Lifelong Learning (CDELL). CDELL is a research centre within the School of Education at the University of Nottingham that undertakes research, development, dissemination and evaluation activities in relation to lifelong learning.

- QCA commissioned CDELL to undertake an independent review of written GCSE examinations in Geography, French and History. The **aim** of the project was to: consider and evaluate the validity (including reliability) and effectiveness of written GCSE examination papers used in these three subjects in order to identify issues that might be addressed in seeking to improve the assessments being made and considering the impact of these on teaching and learning in the subjects concerned.

- The **objectives** of the project were to:

    - review recent reports on GCSEs and relevant research and developments in other countries;
    - examine and evaluate the validity of GCSE examination papers in three subjects, focusing on the range of question types, the use of command words, the effectiveness of coverage of assessment objectives, potential unexpected sources of difficulty, and the use of stimulus material;
    - establish whether the issues identified apply to the subject under investigation or more widely;
    - ascertain how written GCSE examinations might be improved;
    - determine the extent to which potential innovation in GCSE examinations is limited by cost or by the anxiety to ensure acceptable reliability;
    - make recommendations in relation to the future development of written GCSE examination papers and outline where additional research is required.

- The study involved the use of a multi-method approach, which included:

    - a desk review of relevant research developments in the UK and other countries;
    - the use of scrutiny teams, comprising individuals with relevant assessment and subject expertise, to review existing

examination practices and to identify, as a final task, where specific improvements could be made to the GCSE written papers in three subjects in order to improve its validity;

- o a feedback seminar for QCA and awarding body subject officers and/or technical staff to consider and provide feedback on the findings derived from the desk review and scrutinies and to make judgements about which proposals are suitable for further consideration;

- o a conference organised by QCA for awarding body subject officers and technical staff, qualifications development staff and selected senior examiners to consider the implications for GCSE assessment arising from the work.

- The data collected via the scrutiny teams, feedback seminar and conference were analysed and reported at various stages by the CDELL team members, using largely qualitative techniques. The information collected was subjected to an on-going process of data review and progressive refocusing in order to gain a deeper understanding of the respondents' views, experiences and expectations, to develop analytic conclusions, and make broader linkages.

- The main body of this report identifies general, cross-subject improvements that may be made to GCSE assessment. It is based on an analytical review of the principal findings of the scrutiny panels, which can be found in Appendices 1-3, and a wider review carried out by the project team. The subject specific reports in the appendices identify improvements that may be made to GCSE assessment within the three subjects considered, Geography, History and French.

Findings

Predictability

- The structure of many specifications encourages question papers that have predictable formats; this may be unavoidable. Within sections a decrease in the use of stock case study questions and increased overlaps between content areas should reduce predictability. It is important that awarding bodies carefully monitor question papers over the life of a specification to ensure a proper spread of questions.

- With common papers, the structured nature of the questions, whilst providing 'scaffolding' for candidates at the lower end of the target grade range, may prevent those at the top end from demonstrating achievement at the highest level.

Language

The language used in questions

- The multiple uses for the specifications and the diversity of their readerships mean that they must be used with care and not as a blueprint for the language to be used in examination papers.

- Traditional command words are frequently used where simpler, more direct language might better convey the requirement to the candidate. Subtle distinctions between command words do not appear to be understood by candidates, and should not be assumed.

Question presentation

- Questions should be presented in a way that allows candidates to absorb information in a logical and helpful way. Language could in some cases be adjusted accordingly and stock phrases and traditional command words abandoned when simpler and more effective introductions could be used.

- Where it helps candidates to make the transition from one question part to another intermediate stems should be used to clarify the connections and, perhaps, clarify what types of response are needed.

- There is a need to give careful consideration to the demands of specific questions/items. There would be considerable benefits (where possible) in awarding bodies adopting the practice of pre-testing items. Pre-testing would result in greater accuracy in the targeting of items at particular levels/grades.

- There are benefits in there being inclines of difficulty within some GCSE question papers and even within individual structured questions (even though these may not be simple linear progressions). Early parts of such questions should ease candidates in by the use of clear instructions and achievable tasks for the majority. Where it helps candidates to make the transition from one question part to another intermediate stems should be used to clarify the connections and, perhaps, clarify what types of response are needed.

Presentation and Layout

- There is scope for improvement across all of the written papers in the use of text fonts (size, type), highlighting (bold, underlining, italics) and layout (paragraphing and positioning). The use of bold or italic fonts should be kept to an essential minimum and used completely consistently within a paper. Enlargement of fonts may be generally beneficial.

- Graphical and textual source materials should be clearly distinguished from questions, usually through clear placement or boxing. Greater care should be taken in designing the layout of individual pages to ensure that the graphics are positioned in the most appropriate locations in relation to the accompanying text.

- Answer spaces should be set out carefully so as to give clear signals to candidates about the nature and length of answers required. Boxes may sometimes be more appropriate than lines for short or non-text answers.

- Rubrics, source materials, questions and spaces for answers should be so arranged that the candidates do not have to turn over a page in order to link sources to questions.

- There is a need for greater clarity in conveying how the candidates are intended to use the support materials included in written papers. This could be achieved by means of explicit references to all such items in the wording of the questions, and by signposting where they should be used by their positioning in relation to the text.

Designing Questions for All Abilities

- There is sometimes an absence of opportunities in GCSE specifications for candidates at the top of the target grade range to: produce developed arguments, solve problems, make judgments based on critical analysis, and demonstrate mastery of more complex skills.

- A greater emphasis is needed in some cases on questions/sections, which require more analytic and evaluative responses.

- Question rubrics should be suitable for all in the target grade range. This is not always the case at the moment.

Helping Students Understand the Nature of the Challenge

- Candidates are often provided with several indications concerning the nature of the response required: the use of words in the question or stem, the number of marks allocated, the question space and how it is partitioned. These, in turn, relate to the objective and content being assessed. It is important that these indications form a coherent and consistent message. Evidence from examiners' reports reinforces the view that command words do not, by themselves, reliably convey the types of answer required.

- It might be better if levels mark schemes were used only when marking questions eliciting longer, more developed answers than the short responses asked for in the early parts of structured questions.

Impact on Teaching and Learning

- Students' learning experiences could be improved by a reduction in the emphasis in GCSE assessments on the ability to recall small 'chunks' of knowledge and a reduction of 'breadth' of content.

- In some subjects the use of a limited range of predictable question topics and case studies was thought to be having a damaging effect on the students' experience by encouraging an emphasis on teaching small 'chunks' of content and the coaching of examination techniques resulting in superficial coverage of the subject, and a lack of depth and coherence in the students' knowledge and understanding.

- In Geography and History an overemphasis in the specifications on testing of knowledge recall (combined with predictable questions in the written papers) means that subject content can be divided up into small components. These can then be taught as subject content to be memorised for reproduction in the examination, thus encouraging 'surface' at opposed to 'deep' learning – learning which is driven by extrinsic rather than intrinsic motivation. Similar issues arise in French, where the students' learning goals are highly specific, often leading to narrow and uninteresting approaches to the teaching of the subject.

- The emphasis in the specifications on testing the ability to 'recall, select, organise and deploy knowledge' is likely to favour some students at the expense of others, notably those with learning styles better suited to the critical analysis of historical sources and geographical case study materials.

- A broad commitment to the reduction of breadth and an increase in depth of study might be advantageous and could be achieved through:
    - a reduction in the dependence on the absorption of a large amount of factual information ready for use in the examination;
    - specifications that make linkages between content sections more explicit and examination questions that exploit these;
    - setting questions that focus as strongly as possible on understanding and the use of concepts.

Conclusions

- The close scrutiny of the written examinations in three GCSE subjects (French, Geography and History) across all of the five Awarding Bodies provided an effective lens through which to view the workings of the assessment system and to consider its impact on the Key Stage 4 curriculum.

- The study showed that in all three subjects the components of the assessment system (the GCSE criteria, the subject specifications, the written examination papers, the marking schemes and examiners', reports and other forms of feedback to teachers) appear to be in constructive alignment, i.e. there are clear and logical connections between each of the key elements. This is not surprising given the regulatory framework established by QCA, the Quality Assurance mechanisms of the Awarding Bodies (and QCA) plus the intense public scrutiny to which the whole examination system is subjected.

- Within this system, GCSE courses in all subjects are tightly defined - the intended learning outcomes, and the methods by which these will be assessed being clearly spelled out for examiners, teachers and students. As a result of this transparency, teachers and students know exactly what they have to do to get the best possible examination grades. Consequently, they concentrate attention on this when making decisions about what and how to teach, and what and how to learn, which is perfectly understandable given the context of league tables and target setting in which schools operate.

- Within this assessment-led curriculum there are few (if any) rewards to be gained for anyone (examiners, teachers and candidates) deviating from the tried and tested formula – for example by taking risks with innovative methods. Hence, versions of subjects are taught in such a way that their content can be memorised and reproduced in the written examinations, thus encouraging surface, as opposed to deep learning - learning that is driven by extrinsic, rather than intrinsic, motivation. Short-term

objectives (achieving good examination grades) are often
emphasised at the expense of longer-term aims such as nurturing
self-motivation, creativity, knowing how to learn, and developing
transferable, problem solving and critical thinking skills.

- The current GCSE curriculum as a whole is made up of knowledge
compartmentalised into separate subjects, which offers few
opportunities to study the interrelationships between them - it is a
socially-constructed collection code, as opposed to an integrated
code. The upshot is that there are few if any incentives for schools
to reconfigure the curriculum (and to make facilitative changes e.g.
to the timetable) in order to adopt fresh and innovative approaches
to the curriculum, which would enable candidates e.g. to study the
Geography and History of France as well as to learn its language.

- The students' experiences of the GCSE curriculum help to shape
their identities, and to establish their attitudes to learning and their
study habits at a critical stage (14-16) in their life histories. The
current GCSE may not be preparing them all that well for a future
in which they will need to operate as independent and effective
lifelong learners.

The likely impact of the changes proposed in this report

- The changes proposed in this report as a result of CDELL's scrutiny
of the written examinations in three GCSE subjects could work
cumulatively to improve the validity of those examinations and
have a beneficial impact on the learning experiences of students.
They would involve a real shift away from the assessment of
knowledge (and the simple reproduction of learned material)
towards the assessment of the understanding of concepts and
skills. However, they would not radically transform the curriculum
because they would leave existing GCSE syllabi contents largely
intact.

- Whilst changes in examination papers (and even changes in
specifications) are unlikely to have a marked impact on the status
and popularity of Geography, History and French as GCSE subjects
it is important that they:

  o do not detract from the positions occupied by the subjects;
  o offer curriculum continuity from those experienced in the
    subjects at key stage 3;
  o provide an appropriate basis for study at post-16.

- There is no reason to suppose that the changes suggested here
would significantly increase the length of examination papers nor
result in an increased volume of material to be accommodated on
examination room desks.

- Moreover, none of the present proposals are likely to increase examination costs significantly, though question setters might need more time in preparing papers and mark schemes. Some marginal additional costs might be incurred in the initial stages in improving the design and presentation of the written papers.

Wider issues

- Most of the issues identified above relate to fairly modest changes to the format of written examinations in GCSE Geography, History and French. They result from detailed scrutinies of existing GCSE exams in those three subjects, conducted by highly experienced GCE/GCSE examiners. In our discussion with these subject-based examiners, we also touched upon a set of wider issues, which could usefully inform a more general debate about the future shape of GCSE examining at the QCA Conference. These wider issued include:

  o Ways of incorporating modern technologies into GCSE examining to bring it closer to the everyday lives of young people. The French team in particular suggested that text messaging, emails, video conferencing etc. could be used to improve both teaching and assessment. The capturing of student involvement in subject-based work could usefully support school-based assessments, in order to counteract the artificial nature of timed written GCSE examinations of a traditional kind.

  o The point above also relates to a general feeling that GCSE examinations in many subjects are having a 'deadening' impact on teaching and learning. Their traditional approach can constrain teaching and learning and put young people off wanting to study areas of the curriculum, which could under other circumstances be made more appealing.

  o Pressure on schools to do well in national league tables is also having an undesirable impact on GCSE courses. Many schools are under pressure to increase the numbers of students achieving Grades A-C and this is known to lead to very narrow teaching approaches. The formulaic approaches to teaching and learning adopted in many GCSE subjects, exacerbates this problem, as does the practice of entering some students for large numbers of GCSE examinations all at the same sitting.

Section 1   Introduction

Description of the study

This Project was conducted for QCA by the Centre for Developing and Evaluating Lifelong Learning (CDELL). CDELL is a research centre within the School of Education at the University of Nottingham that undertakes research, development, dissemination and evaluation activities in relation to lifelong learning.

QCA commissioned CDELL to undertake an independent review of written GCSE examinations in Geography, French and History. The **aim** of the project was to: consider and evaluate the validity (including reliability) and effectiveness of written GCSE examination papers used in these three subjects in order to identify issues that might be addressed in seeking to improve the assessments being made and considering the impact of these on teaching and learning in the subjects concerned.

The **objectives** of the project were to:

- review recent reports on GCSEs and relevant research and developments in other countries;

- examine and evaluate the validity of GCSE examination papers in three subjects, focusing on the range of question types, the use of command words, the effectiveness of coverage of assessment objectives, potential unexpected sources of difficulty, and the use of stimulus material;

- establish whether the issues identified apply to the subject under investigation or more widely;

- ascertain how written GCSE examinations might be improved;

- determine the extent to which potential innovation in GCSE examinations is limited by cost or by the anxiety to ensure acceptable reliability;

- make recommendations in relation to the future development of written GCSE examination papers and outline where additional research is required.

The study involved the use of a multi-method approach, which included:

- a desk review of relevant research developments in the UK and other countries;

- the use of scrutiny teams, comprising individuals with relevant assessment and subject expertise, to review existing examination practices and to identify, as a final task, where specific improvements could be made to the GCSE written papers in three subjects in order to improve its validity;

- a feedback seminar for QCA and awarding body subject officers and/or technical staff to consider and provide feedback on the findings derived from the desk review and scrutinies and to make judgements about which proposals are suitable for further consideration;

- a conference organised by QCA for awarding body subject officers and technical staff, qualifications development staff and selected senior examiners to consider the implications for GCSE assessment arising from the work.

The data collected via the scrutiny teams, feedback seminar and conference were analysed and reported at various stages by the CDELL team members, using largely qualitative techniques. The information collected was subjected to an on-going process of data review and progressive refocusing in order to gain a deeper understanding of the respondents' views, experiences and expectations, to develop analytic conclusions, and make broader linkages.

All of the work in the project is undertaken within a code of practice that is based on the recommendations of the British Educational Research Association. This has been made available to all those involved in the project; it gives an assurance that individual contributions to the work of the project will remain confidential and that individuals views will not be attributed in any oral or written reports.

The scrutiny process

Examinations studied

GCSE examinations in the subjects included in this project are conducted within the Code of Practice published by the three regulatory authorities (QCA, CCEA and ACCAC). Although details of this Code have changed over the period of the examinations included in this study its principal provisions have remained the same. Examinations in a particular subject are conducted within the provisions of the GCSE Criteria for that subject, also issued by the regulatory bodies. Both the Code and the Criteria are mandatory frameworks within which the examinations in the subject concerned are conducted.

However, the Code and the Criteria deal with all aspects of GCSE while this project is concerned solely with the written papers in the subject concerned. Although there are some references in this report to the impact of these regulatory documents and to coursework the principal focus is on the written papers.

The scrutinies of the written examination papers were all conducted across the five GCSE awarding bodies on England, Wales and Northern Ireland; these are

- Assessment and Qualifications Alliance (AQA)

- Edexcel

- Northern Ireland Council for the Curriculum, Examinations and Assessment (CCEA)

- Oxford, Cambridge and RSA Examinations (OCR)

- Welsh Joint Education Committee (WJEC)

Generally, the principal examination offered by each awarding body was chosen for scrutiny. The examinations for 2002, 2005 and 2006 were included in the study and written papers at all tiers, with all materials linked to these, were included in scrutinies.

The scrutiny teams

Potential members of scrutiny teams were identified with the help of senior awarding body officers. CDELL requested that candidates for inclusion in the teams should be chief or senior examiners with current or very recent experience of GCSE in the subject concerned. The aim was to establish scrutiny teams of four people, each of which would be serviced by two CDELL team members. The aim was to ensure roughly equal representation of the five awarding bodies across all the scrutiny team, with each team having members from four different awarding bodies.

Scrutiny team members were asked to be available for 2 days of individual scrutiny of examination materials and then for a 2-day team session, conducted at the School of Education in the University of Nottingham; the whole of this work was to be completed during August and September 2007.

The scrutiny team for Geography consisted of:

- Jane Cheema (AQA)
- Jon Epstein (Edexcel)
- Justin Woolliscroft (OCR)
- Geraint Williams (WJEC)
- Carole Mallia (CDELL)
- John Wilmut (CDELL)

The scrutiny team for History consisted of:

- Glenn Timms (AQA)
- John Wright (Edexcel)
- Colin Shepherd (OCR)
- Carole Mallia (CDELL)
- Professor Harry Tolley (CDELL)

The scrutiny team for French consisted of:

- Ewen Bird (AQA)
- Steven Crossland (OCR)
- Charlie Henry (CCEA)
- Gareth Roberts (WJEC)
- Roger Murphy (CDELL)
- John Hamer (CDELL)

Materials supplied to scrutiny team members

Scrutiny team members were supplied with background documents describing the project and their role within it:

- The Role of the Scrutiny Panels
- Improving Written GCSE Examinations (a description of the project)
- Instructions for Scrutineers and Meeting Details

They were also supplied with questions to be addressed in relation to each written paper scrutinised. These were grouped within a structured schedule. Multiple copies of this schedule were sent to each member of the scrutiny team.

CDELL members servicing the scrutiny teams were also supplied with two advisory/support documents:

- Two-Day Scrutiny Panel Meetings: Guidelines for CDELL Team Members
- Scrutiny team 2-day meetings: Agenda for the discussion on Day 2

Scrutiny team members were provided with copies of the GCSE Code of Practice and GCSE subject criteria. In general, for each awarding body, they were provided with the same materials relating to the examinations in the chosen subject in each of 2002, 2005 and 2006, namely:

- the specification (where this had changed in the period concerned, more than one specification was included)

- the written examination papers in both higher and foundation tiers; in Geography there were two papers in each tier

- source materials for use with these papers

- the marking schemes relating to these papers

- the grading outcomes for the papers

- the examiners' reports on these examinations

The examinations scrutinised by the Geography panel were:

- AQA   Specification A (3031)
- CCEA   Geography
- Edexcel   Geography A (1312)
- OCR   Geography B (Avery Hill); conducted with WJEC
- WJEC   Geography A (Mainstream) (159)

The examinations scrutinised by the History panel were:

- AQA   History Specification B (3042)
- CCEA   History (G46)
- Edexcel   History Specification A (1334)
- OCR   History Specification B (1937)
- WJEC   History Specification A (165)

The examinations and materials scrutinised by the French panel were:

- AQA   French Specification A (3651)
- CCEA   French Specification
- Edexcel   French Specification (1226)
- OCR   French (1925)
- WJEC   French Specification

There were a few gaps in the provision of materials but these did not have a significant impact on the scrutiny teams' work.

The scrutiny process

The scrutiny process divided into two clear phases.

Phase 1

This was the 2-day individual scrutiny that each team member was asked to do. This involved a study of the examination materials from each awarding body, and in relation to the QCA Code and Criteria. They were asked to use the schedule to identify issues that they thought needed to be addressed in each paper and to suggest changes that might be considered in order to improve the examination. They were invited to take overviews of each aspect of the paper but were told that there would be opportunities for broader discussion during the two-day team scrutiny. The emphasis during this phase was to be on establishing as comprehensive an information base about these papers as was possible in the time available.

It was clear that there was a danger that the volume of material involved in this study would overwhelm members of the scrutiny teams. It was clearly impossible for them to read all of the materials in depth in the time allowed for the work in Phase 1. They were therefore advised to manage their work by:

- allocating a fifth of their 2 days to each awarding body so that all would be considered, though probably incompletely;

- focusing first on the 2006 examination materials and then to look more briefly at the earlier examinations to see whether there were additional issues raised by comparison with 2006;

- using the schedule flexibly, not attempting to answer every question, but selecting issues that appeared to be of the greatest importance and recording comments on these;

- recognising that many comments (such as those on the specification or general comments on question paper style) would relate to more than one examination paper and need not be entered on the schedules more than once.

Phase 2

Members of the scrutiny team were asked to send completed schedules to Nottingham a few days before the team meeting was due to be held. This made it possible for the CDELL team members to conduct a preliminary review of comments and proposals and to prepare more effectively for the team meeting.

The main part of Phase 2 was the team meeting, which fell naturally into two stages.

The first stage, taking a day, consisted of reviewing comments on each of the examinations from all of the team members and recording those, which were agreed to be significant. With the team members' agreement, copies of all members' comments were circulated to the whole team; this made it possible to work more quickly through comments, picking out those which deserved discussion, than would have been possible if each team members had needed to make a contribution on each point. The examination materials were reviewed awarding body by awarding body, thus establishing a secure analysis that would form the basis for discussion on the second day.

One CDELL team member acted as facilitator for this discussion while the other acted as recorder though both contributed to the discussion. In the introduction to this stage of the work team members were reminded of the objectives of the exercise and, in particular, that it was not the purpose of the project to pass judgement on any examination, or to compare examinations from different awarding bodies. The intention was to ensure that there was a secure understanding of the strengths and limitations of current GCSE Geography, History or French examinations so that improvements could be systematically identified. Where team members were involved in a discussion of an examination paper that they had been involved in setting it was hoped that they would take a full part and not feel obliged to defend their work. If they preferred, they were free to withdraw from such discussions; in fact, none did so.

The second stage switched from the analytic process to the more creative consideration of improvements that may be made to GCSE Geography, History or French papers and the potential impact of implementing such improvements. A checklist of key points that had emerged from the analysis on day 1 was used as a starting point for developing proposals and recommendations for improvement and considering the possible impacts of these. Prompted by QCA, the impact on teaching and learning was a principal consideration but other potential impacts were also considered and recorded. The outcomes of these discussions form the basis of the findings and recommendations in the next section of this report.

Because of the exploratory nature of the scrutinies the CDELL team established only outline agendas for these two stages of the work and adapted these to the progress made in the meeting and the advice of the scrutiny team members. The next section includes some comments on the efficiency and completeness of the whole process.

The present report has been written using the schedules produced by the team members during Phase 1 of their work plus the records of the discussion and analysis during the team meeting. The intention is that there shall be a clear audit trail from the individual comments on specific aspects of the examination papers through to the proposed improvements and the perceptions of the likely impact of these. This is necessarily a qualitative (though systematic) process and, given the very tight timeframe for this part of the project, not perhaps as complete as might have been achieved had there been, for example, more time between the phases or more time between the analytic and creative stages of Phase 2 of the team meeting.

General comments

There were considerable similarities amongst the examination papers studied in this exercise. However, the exercise was also constrained by a number of factors.

- Time and cost constraints made it impossible to apply the same levels of scrutiny to papers provided for all three years. The 2006 papers were prioritised and those for 2005 and 2002 were then checked to see whether they raised new issues.

- The criterion for choice of papers for the scrutiny excluded less popular specifications. It would have been useful to include examples of these, particularly in relation to determining the impact on teaching and learning.

- The nature of the chosen methodology led to quite complex discussions about general issues raised through the examples seen by the team members in the course of the first stage of the work. The risk with this is that issues that were identified and suggestions for improvement are necessarily general. Tracking every issue and suggestion back to the specific examples that were used in the discussion is impossible, but every attempt has been made to

constrain these within the scope defined by QCA and to ensure that the location of the supporting evidence can be found in the schedules completed in *Phase 1* of the scrutiny process.

- The discussion would have benefited from a better understanding of the proposed changes to the GCSE criteria for the three subjects, currently under consideration.

- On several occasions the relationship between the written papers and the coursework was an important element in the discussion. Points relating to coursework were provided in the team members' schedules, but these have not been carried into the main body of this report. Consequently, the constraint that confined the discussion to the written papers has resulted in a partial review.

Summary of key points emerging for each examination studied

The schedules returned by the team members before the team meeting included both general and specific comments, awarding body by awarding body. They are bulky and the CDELL team undertook that the information provided in them would not be attributed in any oral or written report. The comments from the team members were consolidated before the start of the team meeting, and by agreement these consolidated versions were circulated and used as the basis for discussion. However, in order to ensure that the materials circulated in advance of the QCA Conference remains manageable these consolidated comments (and the originals from which they were derived) have not been included, though they have all been retained in a project archive for future review/re-examination should the need arise.

The key issues that come out of the scrutiny of all of the papers

As a consequence of the analysis conducted on day 1 of the scrutiny team meeting a range of issues was identified for further discussion on day 2, with the intention of:

- identifying improvements that may be made in GCSE examination papers;
- assessing the impacts that such improvements might have on teaching and learning, management of the examinations, the nature of the assessment and its predictability, the status of Geography, History and French as subjects at GCSE.

Cross-subject issues

The outcomes of the scrutiny panel reviews were analysed by the CDELL team in order to distill out recommendations in relation to general, cross-subject improvements that may be made to GCSE written examination papers. Our recommendations are introduced in Sections 2 and 3.

The separate reports on the Geography, History and French scrutiny panels on which these sections are based are attached as appendices. These subject-specific reports outline the results of the scrutiny panels and identify where specific improvements could be made to the GCSE within particular subjects in order to improve their validity.

Section 2   Findings

This section identifies general, cross-subject improvements that may be made to GCSE assessment. It is based on a review of the principle findings of the scrutiny panels and a further review conducted by the CDELL team. It considers the following issues:

- Predictability
- Language
- Presentation and Layout
- Designing Questions for All Abilities
- Helping Students Understand the Nature of the Challenge
- Impact on Teaching and Learning

Evidence from the scrutiny panel reports is presented in a series of boxes.

Predictability

- The structure of many specifications encourages question papers that have predictable formats; this may be unavoidable. Within sections a decrease in the use of stock case study questions and increased overlaps between content areas should reduce predictability. It is important that awarding bodies carefully monitor question papers over the life of a specification to ensure a proper spread of questions.

- With common papers, the structured nature of the questions, whilst providing 'scaffolding' for candidates at the lower end of the target grade range, may prevent those at the top end from demonstrating achievement at the highest level.

Geography Scrutiny Panel

It was accepted that there was a degree of predictability in all of the papers seen but that much of this stemmed from their arrangement by topic rather than from the choice of questions used in the topic. The invitation to candidates to deploy information from cases that they had studied was a predictable feature. However, in the best examples seen, a variety of different questions were asked about these case studies, reducing predictability. The best examples of extended answer questions cut down the predictability considerably.

It may be the reviser's task to monitor examination papers for predictability. Both formal and less formal methods for checking appeared to be used and are thought to be important. There are expectations of coverage that have to be fulfilled and it is unlikely that any action by awarding bodies will stop teachers from attempting to predict what will appear in an examination. Geography has, in any case, relatively broad specifications that allow examiners reasonable flexibility and reduce predictability.

Reducing predictability through format changes is a risky business since candidates may be dismayed at seeing an unfamiliar layout. Unexpected or unfamiliar source material can offset some predictability but it needs to be chosen carefully (see below). However, there is a very wide range of good material now available to examiners and its use in new ways might reduce predictability whilst improving the quality of assessment. Thus, an OS map might be used to test the usual range of skills whilst also making a contribution as a source for another question on, for example, glacial landscapes.

History Scrutiny Panel

In general, team members found the question topics to be very predictable, but as such a fair reflection of the specifications from which they were derived. In this respect, the question topics chosen were judged to lend themselves to the assessment of the specified content and skills. However, team members were of the view that as a result of their familiarity with the question topics, the candidates would not find them to be particularly interesting or exciting. To counteract this tendency there was a plea for more questions about the experiences of people who lived in the historical periods being studied, coupled with the inclusion of more interesting source materials offering real issues to investigate.

Despite the reservations expressed about the predictability of question topics team members were critical of the occasional instances of what they regarded as questions on 'obscure topics', which they felt would result in candidates being marked for what they did not know rather than for what they did.

Some concerns were voiced about the extent to which setters (and in turn teachers) were taking full advantage of all the opportunities open to them in the specifications. In saying that they recognised that across the specifications, the subject content in many of the options (especially in so-called 'studies in depth') is so limited that there is little scope for varying the topics on which questions are set from one examination to another. Consequently, it is almost inevitable that after a few years the questions become predictable both in terms of the topics they cover and the formats used by the setters. Such predictability was said by the team to be seen not as a vice but virtue by the majority of teachers who are quick to criticise examiners who take the risk of setting questions that depart in any way from the established 'formula' discernible in past examination papers. Finally, the team considered that for the most part, the overall weighting of content and skills within the written papers was appropriate and in line with the requirements of the relevant specifications. However, some concern as expressed about the variations that can occur across the options in the way in which particular aspects of subject content (such as the 'change over time') are tested.

French Scrutiny Panel

Predictability and a degree of repetition are to an extent both inevitable and desirable. The report from one of the awarding bodies, for example, noted that 'the papers contained the familiar range of topics, text styles and test types that are the regular features of this test – no bad thing!' With some justification schools express concern if there are elements within a paper that are perceived to be outside the norm. As with other aspects of assessment, however, there is a balance to be struck. Given the tight parameters within which they are required to operate, examiners have to be extremely skilful if they are to succeed in producing papers that are sufficiently fresh to be interesting and challenging and sufficiently familiar to avoid provoking complaints. In reality the team felt that examiners in French and modern languages generally had less flexibility than examiners in most other subjects, and this could increase the danger of papers being very predictable and possibly, therefore, lessening the validity of them as tests of candidates' knowledge and understanding.

There were concerns that the scrutiny had revealed a lack of variety of question types in many of the papers, and there was a general desire for awarding bodies to find ways of remedying this. Currently, for example, there was commonly an undue reliance on certain types of 'tick box' response items (e.g. matching pictures to words) - particularly with foundation tier papers.

Language

The language used in questions

- The multiple uses for the specifications and the diversity of their readerships mean that they must used with care and not as a blueprint for the language to be used in examination papers.

- Traditional command words are frequently used where simpler, more direct language might better convey the requirement to the candidate. Subtle distinctions between command words do not appear to be understood by candidates, and should not be assumed.

Geography Scrutiny Panel

There are several observations about the language used in the question papers.

The team noted that one method commonly used to distinguish the tiers was the simplification of some language in a Foundation tier paper, when compared with questions (some of which had the same geographic content) at the corresponding Higher tier. The team was not sure whether this was justifiable: weaker performance in geography might not be

associated with weaker language performance. Changes in language when geographic content was the same would make it more difficult to compare performances, particularly when attempting alignment at Grade C.

There were a few instances where the language was thought to more difficult than it need have been. There seems no reason why language should not generally be as simple as possible and the same for both tiers. The team was unhappy that examiners sometimes justified the use of some terminology on the grounds that 'it is in the specification'; specifications are formal documents written primarily for adult use whereas the language of an examination paper should be close to the language in which the subject is learned – the language of the geography classroom. Inaccessibility of language is then less likely to disadvantage some candidates.

However, it is important to distinguish technical terms from common language. Competence in geography must include competence in the use of some parts of the language of geography and this should not be simplified. It would make it much easier for examiners to use suitable language if there were an agreed glossary of essential geographic terms that candidates are expected to know and use. Some specifications either include such glossaries or refer to published glossaries.

A particular problem with the language of the examination papers is the use of command words. On some papers these were extremely repetitive and team members felt (partly from comments in examiners' reports) that some of the subtleties in examiners' uses of command words were not understood by all candidates. When used repeatedly they inhibited good language use, often producing stereotyped instructions and question stems that could often have been presented in a simpler and more accessible way. This bears on the wider issue of question presentation, discussed below.

Finally, questions actually appear in the form of instructions to questions (that is, ending with "?"). This is clearly a common practice, presumably as a result of a common decision. It does lead, in some instances, to unnecessarily complex or abrupt presentations.

History Scrutiny Panel

Several observations were made about the language used in the question papers – though none of them identifying what were regarded as major causes of concern. This means that in general, the language levels used in the written papers across all of the Awarding Bodies were thought to be appropriate to candidates in the target grade range. Sentence clarity and length were both judged to be acceptable and ambiguities and complexities in the use of command words were avoided. However, the need to use easily understandable plain English, and command/instruction words with which the candidates were familiar meant that little variety or interest was detected in the use of language. Similarly, no evidence of biases in usage that would have disadvantaged certain groups was

identified. However, a few examples were noted where the use of a simpler question format would have helped the setters to avoid using long and complicated sentences e.g. by separating instructions relating to the use of sources from the question.

Question presentation

- Questions should be presented in a way that allows candidates to absorb information in a logical and helpful way. Language could in some cases be adjusted and stock phrases and traditional command words abandoned when simpler and more effective introductions could be used.

- Where it helps candidates to make the transition from one question part to another intermediate stems should be used to clarify the connections and, perhaps, clarify what types of response are needed.

- There is a need to give careful consideration to the demands of specific questions/items. There would be considerable benefits (where possible) in awarding bodies adopting the practice of pre-testing items. Pre-testing would result in greater accuracy in the targeting of items at particular levels/grades.

Geography Scrutiny Panel

Questions in these papers are complex, spread over several pages with a large number of parts. In most cases the presentation of the questions appeared to work well, avoiding the well-known pitfalls associated with complex and lengthy question paper layout. However, there were some areas where team members felt that improvement was possible.

There were questions that were introduced in a quite clumsy way, partly (or largely) because of the need to use command words or standard phrases. Thus, for example, a question may begin with the stem "Study Figure x below"; the figure follows and has a title underneath it. Team members felt that candidates' access to the question would be easier if these phrases were linked, such as "Figure x below shows xxxx".

Thereafter, some questions would have benefited from linking phrases (or stems) that enabled candidates to make a smooth transition from one section of a question to the next. In some instances phrases of this kind could be used either to make links between apparently unconnected question parts or to help making it clear what type of answer was required. Whilst it was probably true that some candidates didn't read such linking material and just dived into the questions, its value to others justified its use.

A few examples were found of questions where the presentation seemed positively inhibiting. Examples are questions where the task to be performed was presented before the source upon which the task was

based; questions where items of source material were presented without preamble and questions where no attempt was made to establish linkages from one question part to another. At the same time, the team recognised the difficulties in achieving optimal presentation in all cases and were anxious that improvements in presentation should not result in longer question papers.

- There are benefits in there being inclines of difficulty within some GCSE question papers and even within individual structured questions (though these may not be simple linear progressions). Early parts should ease candidates in by the use of clear instructions and achievable tasks for the majority. Where it helps candidates to make the transition form one question part to another, intermediate stems should be used to clarify the connections and, perhaps, clarify what types of response are needed.

Geography Scrutiny Panel

Because of the arrangement according to topic area there was no general incline of difficulty through question papers. But there were clear inclines through questions, which frequently had many parts. It was reported that such inclines did exist through problem solving and decision-making papers though the team did not review these papers.

Inclines were thought to be essential and that it was important to ease candidates into a question both by the way in which it was and by the nature of its early parts. Inclines were, however, not simple linear structures since it was often necessary to, for example, have a number of basic questions whose information fed into a later part. At all stages it was important that candidates should not require a correct answer to one question part in order to be able to answer the next although marking schemes could adopt the carry-over methods used in mathematics and the sciences. However, most awarding bodies did have certain safeguards against 'double penalty'.

There was some discussion about what made an easy opening question, particularly since the 'define' type of question was reported to cause considerable difficulty for a number of candidates. It is often answered badly because candidates give explanations or examples instead of a definition. The wording of opening questions should be particularly carefully chosen and the subject matter be chosen so as to allow weaker candidates a good opportunity to demonstrate some achievement and to have access to later questions.

History Scrutiny Panel

The team made a number of observations about the choice of question formats in the written papers they examined. In general they found them to be 'fit for purpose' i.e. appropriate to the content and skill being assessed. In particular, they thought that in common papers, the use of structured questions with careful progression through appropriate inclines of difficulty was a suitable format for all students in the target grade range. However, there was some concern that this approach, whilst being helpful to candidates at the lower end of that target grade range, might not be sufficiently challenging for those at the very top.

French Scrutiny Panel

There was a consensus that the papers scrutinised allowed candidates, at both foundation and higher tiers, to demonstrate their knowledge and skills. For the most part this was achieved via suitably graded questions/tasks across the particular paper following an essentially linear incline of difficulty.
Not all papers, however, followed this model. Some adopted a design more akin to a 'peaks and troughs' model, and there was some discussion amongst the team about which was the more appropriate model.

With both models, the team recognised the difficulties inherent in ensuring that the increased level of difficulty between one question and the next was not so great as to prove too daunting – particularly in the earlier parts of a paper where it was important to ease candidates into it. They acknowledged that not all examiners or papers got this right all of the time!

Presentation and Layout

- There is scope for improvement across all of the written papers in the use of text fonts (size, type), highlighting (bold, underlining, italics) and layout (paragraphing and positioning). The use of bold or italic fonts should be kept to an essential minimum and used completely consistently within a paper. Enlargement of fonts may be generally beneficial.

- Graphical and textual source materials should be clearly distinguished from questions, usually through clear placement or boxing. Greater care should be taken in designing the layout of individual pages to ensure that the graphics are positioned in the most appropriate locations in relation to the accompanying text.

- Answer spaces should be set out carefully so as to give clear signals to candidates about the nature and length of answers required. Boxes may sometimes be more appropriate than lines for short or non-text answers.

- Rubrics, source materials, questions and spaces for answers should be so arranged that the candidates do not have to turn over a page in order to link sources to questions.

- There is a need for greater clarity in conveying how the candidates are intended to use the support materials included in written papers. This could be achieved by means of explicit references to all such items in the wording of the questions, and by signposting where they should be used by their positioning in relation to the text.

Geography Scrutiny Panel

There are several issues concerned with the presentation of text and graphics on the examination papers.

An emphasis on a word or phrase through the use of a bold or italic font was sometimes generally over-used or used inconsistently within a single paper. Team members felt that there were instances where the wrong word was emphasised. Poor or over-used emphasis detracts from the value of highlighting key words in a way that will assist candidates. Generally, team members did not favour the use of a different font as a way of emphasising text though there were some cases (particularly within graphics) where this was helpful.

Team members regretted that some fonts were not as large and they would like but recognised that the examination papers could not afford to be much longer than they already are. However, it is possible that fonts could be enlarged by 1 or 2 points without having much effect on paper length though the use of fonts was normally a matter of awarding body rather than subject practice. Expert advice or research might be valuable in making the selection of the most readable font for this type of application.

There were said to be occasional difficulties in reproducing graphics on the coloured paper normally used to distinguish Foundation tier from Higher tier papers. Because this study used photocopied papers the impact of this was difficult to determine.

It was essential that graphical and textual source materials should be clearly distinguished from questions. Most of the papers scrutinised achieved this by placing source material in a box; where this was not done there were occasionally some difficulties in working through the paper with the possibility that candidates might overlook important information.

Where e-marking was used it was essential that candidates wrote answers within the spaces provided. This puts even greater pressure on examiners and awarding body staff to ensure that the answer spaces are well designed. Some answer spaces seemed out of keeping with the question asked; they were sometimes of the wrong length and it seemed usual to

provide lined spaces even when only a one-word or numeric answer was needed - it was not clear why such short answer spaces were not more often boxed. Answer spaces do provide candidates with one type of clue about the required answer – an issue discussed more below in connection with candidates' appreciation of the marks available for a question.

History Scrutiny Panel

With regard to text, it was thought that there was scope for improvement across all of the written papers in the use of fonts (size, type), highlighting (bold, underlining, italics) and text layout (paragraphing and positioning) – matters closely related to issues discussed under other headings.

The team raised a number of issues concerning the overall design of the written papers. In particular, it was thought that the number of options in the papers some of the specifications studied creates a large number of questions, which in turn leads to very long papers. The upshot of this is that there are too many pages for the candidates to sift through under the pressure of examination conditions in order to find the section of the paper, which contains the questions relating to their chosen option. The alternative approach adopted by some of the Awarding Bodies of providing separate question papers (or booklets) for each option was considered to be preferable. A second concern, relates to the layout of the papers. In general, they were seen as being badly designed with rubrics, source materials, questions and spaces for answers often being squashed together, or arranged in such a way that the candidates have to turn over a page in order to link sources to the related questions. The practice of printing the source materials on a separate pull out double-page spread (or fold-out pages) so that it can be opened up and easily studied in conjunction with the relevant questions was commended as an effective way of dealing with this problem.

It was thought that when addressing issues concerned with the design and presentation of the written papers careful thought should be given to the positioning of rubrics, the choice of fonts and the use highlighting - as well as the wording of the information they contain - in order to achieve the maximum clarity for all candidates across the target grade range.

A number of issues were also identified with regard to the presentation of the pages within the written papers. In general, they were found to be reasonably clear without being visually attractive (though the individual pages on one of the papers examined were described as being like 'a maze'). Similarly, the need to turn pages in order to progress from one section of a question to another or to link source material with the related question was generally found to be appropriate. However, instances were identified where the candidates would need to turn the pages repeatedly in order to construct answers to questions based on the source material to which they had been directed – a problem, which could have been obviated in one instance by simply changing the order of the question set. The team members expressed a number of concerns both on the production of the graphics (i.e. the diagrams and other illustrations)

included in the written papers, their use within the questions set, and their positioning in relation to the text. In particular they were critical of the general standard of production as represented by the graphics included in the sample of written papers they reviewed. Whilst acknowledging that the graphics would have looked better in their original form (as opposed to their photocopied versions), they were of the opinion that they could and should be improved. To that end it was suggested that every effort should be made to make these materials look as 'authentic' as possible through the use of colour where appropriate, and by means of high quality 'art work' and reprographics. They were also recommended that these materials should be structured into the questions set i.e. they would chosen, presented and used to serve clearly defined purposes. In turn, this would mean that in designing the layout of individual pages great care would be taken in placing the graphics in the most appropriate locations in relation to the accompanying text, and establishing 'house rules' for where best to place the information relating to the provenance of source materials.

French Scrutiny Panel

The panel suggested the following as ways of improving the French GCSE examinations:

- the use and quality of illustration on papers, both in the provision of icons for the answering of questions and in the inclusion of "contextualising" illustrations which help provide the subject matter of a given exercise thereby giving useful support to candidates. Variation of illustration to include drawings, cartoons, photographs etc. thereby enhancing the attractiveness of presentation;
- variation in the use of fonts including the use of bold type and suitable devices to emphasis key points in rubrics;
- careful consideration of the readability of rubrics and instructions in both English and French.

Designing Questions for All Abilities

- There is sometimes an absence of opportunities in GCSE specifications for candidates at the top of the target grade range to: produce developed arguments, solve problems, make judgments based on critical analysis, and demonstrate mastery of more complex skills.

- A greater emphasis is needed in some cases on questions/sections, which require more analytic and evaluative responses.

- Question rubrics should be suitable for all in the target grade range. This is not always the case at the moment.

Geography Scrutiny Panel

It has generally been difficult to allocate questions with certainty to a 'knowledge' objective rather than and 'understanding' objective and vice versa; there is a considerable grey area between these. The team welcomed the proposal in the draft revised GCSE criteria that places these objectives together and allocates a rather lower percentage weighting to the combined objectives, with a higher weighting to application of knowledge and understanding. It was thought that this would help examiners to set a range of questions that would encourage more attention in the classroom to the learning of concepts, principles and broad skills and that would assess the use of these in both familiar and unfamiliar contexts.

The question papers used in this study were mostly partitioned according to the content area addressed by the question. Almost all questions had a multi-part structure. Only in a few instances did a question introduce items of content from other areas of the specification and there was frequent use of case study questions; these invited the candidate to answer a generic question using information from a topic/area/system that they had previously studied.

Consequently, many of the questions, though well constructed, were limited in the degree to which they explored relationships between different areas of geography and therefore the extent to which they assessed higher order skills. Such skills may require candidates, for example, to assemble relevant evidence, interpret this and evaluate a finding (these are implicit in AO4 in the current GCSE criteria for geography).

The team accepted that the organisation of this study had resulted in the exclusion of specifications that used a different approach. Members reported on a range of alternative awarding body specifications that provided for extended questions and tasks designed to explore knowledge and skills across a wider range of geography; these are of a problem solving type or as a decision making exercise. However, these were not the most popular specifications and there clearly was a majority of teachers who preferred the more 'traditional' types of specification and question paper.

The difficulty with the widespread use of case studies is that candidates may learn these by heart and present these in answer to generic questions without having to deploy much in the way of reasoning, analysis or evaluation. It is extremely difficult for examiners to ensure, through the wording of generic questions, that these skills are deployed; consequently, much of an examination paper is reduced to the assessment of recall. This has knock-on effects on learning: there is a perception that geography is very demanding because of the amount to be learned and teachers are discouraged from departing from a fixed menu of case studies. This discourages innovative teaching approaches in, for example, switching to the study of a current phenomenon (a hurricane or the local discussion of a proposed shopping mall) to which students

might readily relate.

The team saw no simple solution to this problem. It would not be appropriate to either abandon the use of case studies or require that all students took a problem-solving or decision-making paper. However, it was possible to limit the use of case study responses and to replace some of these with questions based on scenarios, forcing candidates to apply principles or offer appropriate examples or solutions that are drawn from their own experience but applied to the given scenario.

The team felt that multi-part questions should generally build towards sections that required more analytic responses and, where possible, drew information from across the specification and perhaps analysed or evaluated it, rather than confining it to one content section.

Practices with problem-solving or decision-making papers were said to vary: some awarding bodies sent out materials in advance but others did not. Some created whole papers of this type while others used only a part of a paper for this purpose. There were clear and obvious cases for both approaches and the team was anxious to preserve as much choice as possible for teachers. In using materials of this kind, especially where these might be created around the assessment of concepts, it was important to distinguish between what was appropriate for GCSE and what should be reserved for A level or other Level 3 programmes.

The examination papers that were studied offered a wide range of source material that enabled candidates to deploy their knowledge and skills. There were occasional questions that team members criticised but the general view was that the quality of the papers was high.

The scrutiny of some questions triggered a discussion of the extent to which it was possible or legitimate to set questions that used wholly unfamiliar settings within which candidates could deploy their knowledge and skills. There was one particular question which sought to do this but where the choice of setting was judged to be too far from the candidates' experience, thus inhibiting good responses by presenting an unnecessary barrier. A little more explanation or the choice of a slightly more familiar setting would have greatly improved the question.

However, it was felt that moving candidates into unfamiliar territory was desirable, as it forces them to apply their geographical knowledge and skills in a different context to that which they have learnt, thus getting away from the regurgitation of previously learned case study material with its dependence on recall. Questions that invite candidates to deploy ideas, comparisons or information from cases that they have studied, but applying it to an unfamiliar setting, are likely to be more satisfactory. If such questions then require candidates to synthesise and evaluate information that has come from several sources then they provide a more satisfactory assessment of AO3 and AO4.

The team recognised that there should be a limit on the settings that should be used (they must not stray too far from familiar contexts) and

that it remains important to assess knowledge and understanding.

History Scrutiny Panel

The concerns of the team with the number of option choices candidates were faced with in some of the papers they were asked to study were reflected in the issues they raised concerning general rubrics. By and large they thought that the instructions to candidates set out in the general rubrics located on the front of question papers were too complicated - or 'challenging' for those at the lower end of the target grade range. By comparison, localised rubrics were generally seen as being clear and straightforward.

The team made a number of observations about the choice of question formats in the written papers they examined. In general they found them to be 'fit for purpose' i.e. appropriate to the content and skill being assessed. In particular, they thought that in common papers, the use of structured questions with careful progression through appropriate inclines of difficulty was a suitable format for all students in the target grade range. However, there was some concern that this approach, whilst being helpful to candidates at the lower end of that target grade range, might not be sufficiently challenging for those at the very top. This was coupled with the view that the form of response invited by the questions set, was appropriate for testing recall and description (for allowing the candidates to 'regurgitate what they have been taught') – but less so for demonstrating higher order skills, which need more thought-provoking questions. Perhaps because the team are all seasoned examiners, and as such are all accustomed to seeing question papers at all stages in their development and use, they found the papers to be both predictable and unexciting.

By and large, the team members declared themselves to be 'fairly confident' that the specifications they had studied from across the Awarding Bodies would lead to valid and reliable grade outcomes. There were however, a number of reservations. The first related to what was seen as the absence of opportunities for candidates at the top end of the target grade range (i.e. A/A*) to demonstrate their ability to produce developed arguments, solve problems and make judgments based on critical analysis, as opposed to reproducing what they have learned (or 'jumping through hoops without thinking'. The second related to a concern about parity in the levels of difficulty between the different option papers open to candidates. In this respect, it was noted that whilst the question set can be broadly comparable in terms of their difficulty, differences often exist in the related source materials making some options appear at the surface to be more challenging for the candidates than others.

Helping Students Understand the Nature of the Challenge

- Candidates are often provided with several indications concerning the nature of the response required: the use of words in the question or stem, the number of marks allocated, the question space and how it is partitioned. These, in turn, relate to the objective and content being assessed. It is important that these indications form a coherent and consistent message. Evidence from examiners' reports reinforces the view that command words do not, by themselves, reliably convey the types of answer required.

Geography Scrutiny Panel

Command words have already been discussed in connection with the language and text of the question paper. However, these words are traditionally seen to provide information about the type of answer that is required, so that the use of a word like 'define' points to the need for a different type of response than one that would follow from the use of a word like 'explain' or 'describe'. Teachers have had the task of preparing candidates to respond in appropriate ways to these words but it was noticeable that many examiners' reports included comments that suggested that candidates did not produce the type of answer that was indicated by the command word that was used. Of course, in many cases this was probably because they were unable to do so but there is the nagging concern that over-reliance on these words and the subtle shades of meaning that they may convey is lost on some candidates who, given more explicit information, could produce a more creditable response.

A general simplification of language, bringing it closer to classroom language use, and the use of more explicit instructions that identify what is required from an answer could reduce the dependence on this stock of command words. Stems that make linkages between question parts and careful management of the answer spaces, both already discussed, may help. The whole should be explicit and not dependent on assumptions about the way in which teachers have prepared candidates for the examination.

Clarification of this type should be seen in the context of what is being assessed. It is not being suggested that questions are broken down into simplified bits when it is the candidate's capacity to assemble and present an argument or to evaluate some information that is being assessed. However, the wording of such questions can often point the candidate more clearly in the direction of the answer that is required without such a heavy dependence on stock command words.

Candidates already know how many marks are available for a question part, and this provides a further clue about the answer required. The team members felt that marking schemes (particularly when levels were being used) occasionally failed to relate the rewards to the scope and depth of answers as described in the question being asked. It is thought doubtful whether it helps candidates to know which questions are level marked

though many of them are probably aware of the existence of levels marking and that it is likely to be used for longer questions.

- It might be better if levels mark schemes were only used when marking questions eliciting longer more developed answers than the short responses asked for in the early parts of structured questions.

Geography Scrutiny Panel

Team members were concerned that levels marking was sometimes used for questions attracting only 2 or 3 marks. It was thought that levels marking should only be used when questions attract 5 or more marks, perhaps occasionally 4.

Markers need to be provided with clear triggers that enable them to distinguish levels before making a judgment within a level. A trigger will often be in terms of the type of skill being deployed in the answer. It would be good practice to include small examples of level marked work in mark schemes, but ensuring that markers understand that these are not the only acceptable answers.

The status of levels appeared differ amongst awarding bodies. In some, all three levels were used in both tiers whereas elsewhere three levels were used for the Higher tier and two for the Foundation tier though at least one awarding body was not rigid about this. In the latter case there appeared to be some notional relationship between grades and levels: level 3 might correspond to performances typical of grades B, A and A*. This might not matter unless, in lieu of making the levels scheme explicit, markers were briefed in terms that described a level by reference to what was expected of a candidate with a typical grade performance. That might fail to reward answers that were atypically good or poor.

However, the team members recognised that awarding bodies may provide general briefing that was not available to the scrutiny and that might clarify this issue. Without access to fuller information it was not being suggested that all awarding bodies should use levels in exactly the same way (they can simply be regarded as units of account within the marking process).

It is common for question papers to include some anchor questions that are common to both tiers; these both reflect the fact that the tiered papers are assessing on the same specification and that it is helpful to have such questions when considering the relationship between the papers, particularly, the overlap at Grade C.

Some awarding bodies had up to 40% of the questions in common; others had less, perhaps on the grounds that this detracted from the value of having tiers. Elsewhere, questions on the tiers were comparable, though with detailed differences.

Equating at Grade C seems likely to be easiest if the anchor questions are

identical on both tiers and the marks awarded are the same. Here it will be possible to introduce statistical evidence of performance that can fairly easily be interpreted alongside expert judgments. The equating probably becomes a little more awkward (and a little more dependent on judgment) if the marks awarded are different but on parallel linear scales and/or the Foundation tier question is the same as the Higher tier question but with, say, one part removed. Equating seems to get significantly more difficult if, say, the questions have identical content but different wording and most difficult if there are very limited or no anchor questions. In these cases any statistical evidence will be very difficult to interpret.

It is difficult to see why question wording should be different on both tiers if it the same content and skills being assessed. Wording that is suitable for Foundation tier candidates should also be used for Higher tier candidates.

In general, a higher proportion of common questions seem more appropriate than a lower proportion and, except where it is appropriate to remove some question parts for Foundation tier candidates, the questions should be identical with identical marking schemes. This does suggest that there may be occasions where level 3 may be used for Foundation tier.

History Scrutiny Panel

By and large, the team members found the marking schemes they examined to be clearly presented, supportive of the marking process and allowing the knowledge, understanding and skills demonstrated by the candidates to be appropriately rewarded. That said some observations were made about 'levels mark schemes'; an approach, which seemed to the team to be more appropriate for the marking of questions demanding longer more developed answers than to the short responses required in the early parts of structured questions carrying few marks. Whilst commending the appropriate use of this method of marking, the team noted that some of the descriptors were 'too rambling' making it difficult in their opinion, to identify distinct levels of performance. It was pointed out some of the marking schemes the guidance looked 'too thin' and was lacking in examples. This was counteracted in other cases where it was said that the marking schemes are beginning to look like 'notes for answers' – an attempt, it was suggested to provide guidance for assistant examiners who are themselves uncertain about knowledge and understanding of particular content options. Finally, the team voiced some reservations about the use of marking schemes outside of the examination context for which they were devised, and for purposes other than those for which they were originally intended i.e. to facilitate consistent and reliable marking. As such they are working documents for use in standardising meetings (at which the principles underpinning them, and their practical applications are clarified by senior examiners) and in the subsequent marking of scripts. In particular, there was concern over their widespread use for coaching candidates for the written examinations, and the way they deflect interest away from consideration of wider curriculum and assessment issues at INSET meetings.

French Scrutiny Panel

All marking schemes should include a combination of *general* marking guidelines and *specific examples* taken from candidates' work on the current examination. Schemes as given to examiners should be reproduced fully in the report to Centres. In addition, students should always be shown how marks are apportioned within different parts of written examination papers.

Impact on Teaching and Learning

- Students' leaning experiences could be improved by a reduction in the emphasis in GCSE assessments on the ability to recall small 'chunks' of knowledge and a reduction of 'breadth' of content.

- In some subjects the use of a limited range of predictable question topics and case studies was thought to be having a damaging effect on the students' experience by encouraging an emphasis on teaching small 'chunks' of content and the coaching of examination techniques resulting in superficial coverage of the subject, and a lack of depth and coherence in the students' knowledge and understanding.

- In Geography and History an overemphasis in the specifications on testing of knowledge recall (combined with predictable questions in the written papers) means that subject content can be divided up into small components. These can then be taught as subject content to be memorised for reproduction in the examination, thus encouraging 'surface' at opposed to 'deep' learning – learning which is driven by extrinsic rather than intrinsic motivation. Similar issues arise in French, where the students' learning goals are highly specific, often leading to narrow and uninteresting approaches to the teaching of the subject.

- The emphasis in the specifications on testing the ability to 'recall, select, organise and deploy knowledge' is likely to favour some students at the expense of others, notably those with learning styles better suited to the critical analysis of historical sources and geographical case study materials.

- A broad commitment to the reduction of breadth and an increase in depth of study would be welcome and might be achieved through:

  o a reduction in the dependence on the absorption of a large amount of factual information ready for use in the examination;

  o specifications that make linkages between content sections more explicit and examination questions that exploit these;

  o setting questions that focus as strongly as possible on understanding and the use of concepts.

Section 3   Conclusions

- The close scrutiny of the written examinations in three GCSE subjects (French, Geography and History) across all of the five Awarding Bodies provided an effective lens through which to view the workings of the assessment system and to consider its impact on the Key Stage 4 curriculum.

- The study showed that in all three subjects the components of the assessment system (the GCSE criteria, the subject specifications, the written examination papers, the marking schemes and examiners', reports and other forms of feedback to teachers) appear to be in constructive alignment, i.e. there are clear and logical connections between each of the key elements. This is not surprising given the regulatory framework established by QCA, the Quality Assurance mechanisms of the Awarding Bodies (and QCA) plus the intense public scrutiny to which the whole examination system is subjected.

- Within this system, GCSE courses in all subjects are tightly defined - the intended learning outcomes, and the methods by which these will be assessed being clearly spelled out for examiners, teachers and students. As a result of this transparency, teachers and students know exactly what they have to do to get the best possible examination grades. Consequently, they concentrate attention on this when making decisions about what and how to teach, and what and how to learn, which is perfectly understandable given the context of league tables and target setting in which schools operate.

- Within this assessment-led curriculum there are few (if any) rewards to be gained for anyone (examiners, teachers and candidates) deviating from the tried and tested formula – for example by taking risks with innovative methods. Hence, versions of subjects are taught in such a way that their content can be memorised and reproduced in the written examinations, thus encouraging surface, as opposed to deep learning - learning that is driven by extrinsic, rather than intrinsic, motivation. Short-term objectives (achieving good examination grades) are often emphasised at the expense of longer-term aims such as nurturing self-motivation, creativity, knowing how to learn, and developing transferable, problem solving and critical thinking skills.

- The current GCSE curriculum as a whole is made up of knowledge compartmentalised into separate subjects, which offers few opportunities to study the interrelationships between them - it is a socially-constructed collection code, as opposed to an integrated code. The upshot is that there are few if any incentives for schools to reconfigure the curriculum (and to make facilitative changes e.g. to the timetable) in order to adopt fresh and innovative approaches

to the curriculum, which would enable candidates e.g. to study the Geography and History of France as well as to learn its language.

- The students' experiences of the GCSE curriculum help to shape their identities, and to establish their attitudes to learning and their study habits at a critical stage (14-16) in their life histories. The current GCSE may not be preparing them all that well for a future in which they will need to operate as independent and effective lifelong learners.

The likely impact of the changes proposed in this report

- The changes proposed in this report as a result of CDELL's scrutiny of the written examinations in three GCSE subjects could work cumulatively to improve the validity of those examinations and have a beneficial impact on the learning experiences of students. They would involve a real shift away from the assessment of knowledge (and the simple reproduction of learned material) towards the assessment of the understanding of concepts and skills. However, they would not radically transform the curriculum because they would leave existing GCSE syllabi contents largely intact.

- Whilst changes in examination papers (and even changes in specifications) are unlikely to have a marked impact on the status and popularity of Geography, History and French as GCSE subjects it is important that they:

  - o do not detract from the positions occupied by the subjects;
  - o offer curriculum continuity from those experienced in the subjects at key stage 3;
  - o provide an appropriate basis for study at post-16.

- There is no reason to suppose that the changes suggested here would significantly increase the length of examination papers nor result in an increased volume of material to be accommodated on examination room desks.

- Moreover, none of the present proposals are likely to increase examination costs significantly, though question setters might need more time in preparing papers and mark schemes. Some marginal additional costs might be incurred in the initial stages in improving the design and presentation of the written papers.

Wider issues

- Most of the issues identified above relate to fairly modest changes to the format of written examinations in GCSE Geography, History and French. They result from detailed scrutinies of existing GCSE exams in those three subjects, conducted by highly experienced GCE/GCSE examiners. In our discussion with these subject-based examiners, we also touched upon a set of wider issues, which could

usefully inform a more general debate about the future shape of GCSE examining at the QCA Conference. These wider issued include:

o Ways of incorporating modern technologies into GCSE examining to bring it closer to the everyday lives of young people. The French team in particular suggested that text messaging, emails, video conferencing etc. could be used to improve both teaching and assessment. The capturing of student involvement in subject-based work could usefully support school-based assessments, in order to counteract the artificial nature of timed written GCSE examinations of a traditional kind.

o The point above also relates to a general feeling that GCSE examinations in many subjects are having a 'deadening' impact on teaching and learning. Their traditional approach can constrain teaching and learning and put young people off wanting to study areas of the curriculum, which could under other circumstances be made more appealing.

o Pressure on schools to do well in national league tables is also having an undesirable impact on GCSE courses. Many schools are under pressure to increase the numbers of students achieving Grades A-C and this is known to lead to very narrow teaching approaches. The formulaic approaches to teaching and learning adopted in many GCSE subjects, exacerbates this problem, as does the practice of entering some students for large numbers of GCSE examinations all at the same sitting.

Appendix 1

Geography Scrutiny Exercise

This report identifies improvements that may be made to GCSE assessment within Geography. It is based on a review of the principle findings of the Geography scrutiny panel. It considers the following issues:

- Specifications in relation to the GCSE criteria
- Written paper content
- Written paper presentation
- Marking issues

Follow a summary of the key issues and recommendations for improvement, and a discussion of the impacts of these, evidence from the Geography scrutiny panel report is presented in a series of boxes numbered G1-G16. Links between the key issues/recommendations and the more detailed evidence contained in the boxes are clearly identified.

The Geography scrutiny panel meeting was conducted on 20-21 August.

The scrutiny team for Geography consisted of

Jane Cheema (AQA)
Jon Epstein (Edexcel)
Justin Woolliscroft (OCR)
Geraint Williams (WJEC)
Carole Mallia (CDELL)
John Wilmut (CDELL)

Scrutiny outcomes

Issues and recommendations for improvement

1   Specifications in relation to the GCSE criteria

1.1   There are many differences between the specifications from the different awarding bodies. *[See Box G1 below]*

1.2   The multiple uses for the specification and the diversity of its readership mean that it must used with care and not as a blueprint for the language to be used in examination papers. *[See Box G1 below]*

1.3   The proposed merging, in the revised GCSE criteria, of knowledge and understanding into a single objective was welcomed (a) as a simplification that would assist examiners and (b) as a contribution to a shift towards the better learning of geographical concepts and principles. *[See Box G2 below]*

1.4   It was hoped that some diversity would be maintained between specifications, giving teachers choice and allowing for a variety of

approaches to teaching and learning in geography. *[See Box G3 below]*

1.5   The limitations of an approach based solely on candidates' reproduction of learned case studies should be recognised. *[See Box G3 below]*

1.6   Geography appears to be very demanding because, for some students, it involves the absorption of a large amount of factual information, ready for use in the examination. Teachers may be reluctant to depart from a fixed menu of case studies and are reluctant to use contemporary material that will enable students to develop concepts whilst feeling that they are dealing with relevant material. *[See Boxes G3 and G6 below]*

1.7   Where possible, questions should use information and concepts from across the specification and/or allow students to draw on information from more than one case study. Questions should normally build towards sections that require more analytic and evaluative responses. *[See Boxes G3 and G9 below]*

1.8   The award of marks for QWC should be made explicit in marking schemes and should generate comparable outcomes whether manual or e-marking is used, and across boards. *[See Box G4 below]*

1.9   Percentage weightings for objectives should be used similarly by the awarding bodies so that reasonable consistency can be maintained between years. *[See Box G5 below]*

1.10   A broad commitment to the reduction of breadth and an increase in depth of study would be welcome and might be achieved through a reduction in the dependence on case studies (particularly those that appear in standard publications), encouragement to teachers to develop and use contemporary case study materials, specifications that make linkages between content sections more explicit and examination questions that exploit these, with questions that focus as strongly as possible on the understanding and use of concepts. *[See Boxes G6 and G9 below]*

2   Written paper content

2.1   Setting some questions that present candidates with unfamiliar settings that demand that they use their understanding of geographical concepts, deploying ideas, making comparisons or providing information from cases that they have studied, are likely to reduce the focus on simple recall. *[See Box G7 below]*

2.2   It remains, however, essential that the required amount of knowledge and understanding is assessed and that settings are not so unfamiliar that candidates cannot cope with them. *[See Box G7 below]*

2.3   It is important that there are inclines of difficulty within questions (though these may not be simple linear progressions). Early parts should ease candidates into the question by the use of clear

instructions and achievable tasks for the majority. *[See Box G8 below]*

2.4 The structure of many specifications encourages question papers that have predictable formats; this may be unavoidable. Within sections a decrease in the use of stock case study questions and increased overlaps between content areas should reduce predictability. It is important that awarding bodies carefully monitor question papers over the life of a specification to ensure a proper spread of questions. *[See Box G9 below]*

3 Written paper presentation

3.1 The use of an answer booklet may place additional demands on candidates. *[See Box G10 below]*

3.2 The use of bold or italic fonts should be kept to an essential minimum and used completely consistently within a paper. *[See Box G11 below]*

3.3 Enlargement of fonts may be generally beneficial. Otherwise, research or expert advice on fonts would be welcome. *[See Box G11 below]*

3.4 Graphical and textual source materials should be clearly distinguished from questions, usually through clear placement or boxing. *[See Box G11 below]*

3.5 Answer spaces should be managed carefully so as to give clear clues to candidates about the nature of answers required. Boxes may sometimes be more appropriate than lines for short or non-text answers. *[See Box G11 below]*

3.6 The simplification of Foundation tier paper language compared with higher tier may not be justified; it is usually more appropriate to use common language, simplifying the higher tier paper. *[See Boxes G12 and G16 below]*

3.7 It is not an adequate test of language level to say that 'it is in the specification'. Language should generally reflect usage in the geography classroom. An exception is the use of words that are specific to geography (technical words) that should be distinguished in a glossary. *[See Box G12 below]*

3.8 Traditional command words are frequently used where simpler, more direct language would better convey the requirement to the candidate. Subtle distinctions between command words do not appear to be understood by candidates, and should not be assumed. *[See Box G12 below]*

3.9 Questions should be presented in a way that allows candidates to absorb information in a logical and helpful way. Language should be adjusted accordingly and stock phrases and traditional command words abandoned when simpler and more effective introductions can be used. *[See Box G13 below]*

3.10 Where it helps candidates to make the transition from one question part to another intermediate stems should be used to clarify the connections and, perhaps, clarify what types of response are needed. *[See Box G13 below]*

4   Marking issues

4.1   Evidence from examiners' reports reinforces the view that command words do not, by themselves, reliably convey the types of answer required. *[See Box G14 below]*

4.2   Candidates are provided with several clues about the nature of the response required: the use of words in the question or stem, the number of marks allocated, the question space and how it is partitioned. These, in turn, relate to the objective and content being assessed. It is important that these clues form a coherent and consistent message. *[See Box G14 below]*

4.3   Levels marking should only be used for longer question parts. *[See Box G15 below]*

4.4   There should be clear triggers that enable markers to distinguish between levels. Examples could usefully be used to illustrate differences between levels and differences within levels, helping markers to make consistent judgments. *[See Box G15 below]*

4.5   Consistency amongst awarding bodies in the use of levels may be helpful. Linking levels to notional grades carries the risk that markers may substitute a perception of a grade for the definition of the level. *[See Boxes G15 and G16 below]*

4.6   It is easier and perhaps more satisfactory to equate Grade C performances if there are larger amounts of common questions on the tiered papers. Changing the language between tiers but not the content of questions is unsatisfactory. *[See Box G16 below]*

Reviewing impacts

5   Impacts on Teaching and Learning

5.1   The amount of content in the specifications has an effect on the take-up of geography and on the ways in which it is taught. Reductions in this with a compatible use of examination questions that exploit connections across the specification should have a beneficial impact.

5.2   Whilst students will need to continue to be taught good examination techniques, a reduction in the use of command words, a general clarification of language and improved presentations should serve to emphasise the geographic content rather than the mechanics of the examination.

5.3   The presence of questions that draw on material from across the specification, perhaps linked to candidates' own case studies should

discourage the widespread learning of case study material that is then recalled as stock answers to generic examination questions.

5.4   A reduction in the breadth of content in some specifications, linked to a greater use of cross-specification questions may free teachers to use more contemporary case studies that should have a greater appeal to students.

5.5   The use of unfamiliar application questions will have an effect on how the subject is taught, enabling teachers to deploy materials and students to research sources, so developing concepts that they will then be able to apply in different settings.

5.6   The use of decision-making and problem solving tasks (as whole or part-papers) places a very direct focus on the development of understanding and skills rather than the learning of stock material. It will, however, need more support from the awarding bodies.

6   Impact on the management and administration of examinations

6.1   It is essential that the types of change suggested here do not significantly increase the length of examination papers nor result in an increased volume of material to be accommodated on examination room desks. There is no reason to suppose that it will.

7   Impact on the cost of the examinations

7.1   None of the present proposals are likely to increase examination costs significantly though question setters might need more time in preparing papers and mark schemes.

8   Impact on the range of content and skills being assessed

8.1   The proposals here envisage a real shift away from the assessment of knowledge (and the simple reproduction of learned material) towards the assessment of concepts and skills.

8.2   It is essential that examination papers give as rewarding an experience as possible to lower achieving candidates whilst ensuring high rewards for high achieving candidates.

9   Impact on the status and popularity of the subject

9.1 Whilst changes in examination papers (and even changes in specifications) are unlikely to have a marked impact on the status and popularity of geography as a GCSE subject it is important that they

a) do not detract from its position

b) represent a proper continuity from geography at key stage 3

c) offer an appropriate basis for study post-16

Evidence

The issues are discussed in more detail in the boxes below in an order that follows that of the scrutiny schedule - though in the discussion itself members of the team frequently linked issues in ways that were more holistic and discussions and cut across the schedule's structure. For each of the issues there is a discussion in the left-hand column and, in the right, a cross-reference to a proposal or recommendation from the list in the section above.

Specifications in relation to the GCSE criteria

Box G1: Presentation of the specifications	
A specification is a very complex statement of intent on the part of the awarding body that generates it. Within the statutory and regulatory framework it identifies the nature of the qualification to be awarded in terms of its content, assessment arrangements and performance requirements. It is a public document addressed to all those concerned with the GCSE in the subject but is particularly directed at and important to teachers (who will interpret it to their students) and examiners (who use it as the framework for setting and marking the examination). Alongside the GCSE Code of Practice and the GCSE criteria for geography, it defines the basis for determining the validity of the grades awarded on the examination. The team members made a distinction between the use of the specification to define, for examiners, moderators, revisers and others, the scope of the examination and its use by teachers who use it as a curriculum-planning guide and interpret its contents to students. They were surprised at a number of differences between the specifications from the five awarding bodies (discussed below) and concerned that the language used in the specifications should not provide blueprints for the language of the examination papers (also see below).	**Issues and recommend-ations 1.1 and 1.2**

Box G2: Knowledge & understanding	
It has generally been difficult to allocate questions with certainty to a 'knowledge' objective rather than an 'understanding' objective and vice versa; there is a considerable grey area between these. The team welcomed the proposal in the draft revised GCSE criteria that places these objectives together and allocates a rather lower percentage weighting to the combined objectives, with a higher weighting to application of knowledge and understanding. It was thought that this would help examiners to set a range of questions that would encourage more attention in the classroom to the learning of concepts, principles and broad skills and that would assess the use of these in both familiar and unfamiliar contexts.	**Issues and recommend-ations 1.3**

Box G3: Synthesis	
The question papers used in this study were mostly partitioned according to the content area addressed by the question. Almost all questions had a multi-part structure. Only in a few instances did a question introduce items of content from other areas of the specification and there was frequent use of case study questions; these invited the candidate to answer a generic question using information from a topic/area/system that they had previously studied. Consequently, many of the questions, though well constructed, were limited in the degree to which they explored relationships between different areas of geography and therefore the extent to which they assessed higher order skills. Such skills may require candidates, for example, to assemble relevant evidence, interpret this and evaluate a finding (these are implicit in AO4 in the current GCSE criteria for geography). The team accepted that the organisation of this study had resulted in the exclusion of specifications that used a different approach. Members reported on a range of alternative awarding body specifications that provided for extended questions and tasks designed to explore knowledge and skills across a wider range of geography; these are of a problem solving type or as a decision making exercise. However, these were not the most popular specifications and there clearly was a majority of teachers who preferred the more 'traditional' types of specification and question paper. The difficulty with the widespread use of case studies is that candidates may learn these by heart and present these in answer to generic questions without having to deploy much in the way of reasoning, analysis or evaluation. It is extremely difficult for examiners to ensure, through the wording of generic questions, that these skills are deployed;	**Issues and recommend-ations 1.4 to 1.7**

consequently, much of an examination paper is reduced to the assessment of recall. This has knock-on effects on learning: there is a perception that geography is very demanding because of the amount to be learned and teachers are discouraged from departing from a fixed menu of case studies. This discourages innovative teaching approaches in, for example, switching to the study of a current phenomenon (a hurricane or the local discussion of a proposed shopping mall) to which students might readily relate.

The team saw no simple solution to this problem. It would not be appropriate to either abandon the use of case studies or require that all students took a problem-solving or decision-making paper. However, it was possible to limit the use of case study responses and to replace some of these with questions based on scenarios, forcing candidates to apply principles or offer appropriate examples or solutions that are drawn from their own experience but applied to the given scenario.

The team felt that multi-part questions should generally build towards sections that required more analytic responses and, where possible, drew information from across the specification and perhaps analysed or evaluated it, rather than confining it to one content section.

Practices with problem-solving or decision-making papers were said to vary: some awarding bodies sent out materials in advance but others did not. Some created whole papers of this type while others used only a part of a paper for this purpose. There were clear and obvious cases for both approaches and the team was anxious to preserve as much choice as possible for teachers. In using materials of this kind, especially where these might be created around the assessment of concepts, it was important to distinguish between what was appropriate for GCSE and what should be reserved for A level or other Level 3 programmes.

Box G4: Quality of written communication (QWC)	
The scrutiny (particularly that of the marking schemes) revealed some differences in the ways in which QWC was rewarded and the ways in which an assessment of QWC would influence a question mark. The team members didn't know whether there were overarching awarding body policies in this matter. It also wasn't clear whether the determination of QWC was a global judgment or a judgment made on a question-by-question basis nor whether the judgment about QWC might, for example, move the reward for an answer from one level to another or only between marks within a level. In the absence of specific guidance, e-marking of some papers (when a marker may not see the whole of a script) might produce a different QWC outcome from that where a marker made a	**Issues and recommend-ations 1.8**

judgment across the whole of a script. Some levels descriptions implicitly or explicitly incorporated QWC in the description of an appropriate answer, but others did not. Neither the GCSE criteria nor the Code of Practice offered much help in this matter. It appears unsatisfactory that the commitment in principle to the assessment of QWC cannot be tracked through to identifiable, consistent and clear-cut practices in script marking although team members acknowledged that those methods to be used might have been presented in documents not available to this scrutiny.	

Box G5: Weighting ranges	
The GCSE criteria show percentage ranges for each of the assessment objectives. Four of the five awarding bodies showed specific weightings for these objectives in their specifications; one awarding body showed quite wide weighting ranges for each objective. All weightings were within the ranges in the criteria. One effect of these ranges within a specification may be to allow more annual variation in the balance of the objectives than would occur if no ranges were specified. Of course, the weightings are, in practice, only achieved approximately since it is very difficult to be sure what is being tested by any question or question part.	**Issues and recommend-ations 1.9**

Box G6: Geography is perceived to be difficult	
Team members were concerned that entries for geography are declining. The two principal reasons were thought to be the availability of subjects such as Leisure and Tourism (that earns 2 GCSE grades) and the perception that geography is a difficult subject; students feel that there is a large amount of content to be learned. Reducing the commitment to learning a range of case studies (already discussed) would go some way to addressing this perception. However, this should be seen as one part of a wider commitment to the study of the subject in more depth, at the expense of some of the current breadth. This might include greater opportunities for using examples from the present-day, as the occasion arises, using internet and other contemporary material, together with a rather stronger focus on developing the understanding and use of concepts. In support of this it was essential that examination questions provided students with opportunities to introduce material from a range of courses and that the dependence on textbook sources was reduced.	**Issues and recommend-ations 1.6 and 1.10**

Written paper content

Box G7: Familiar and unfamiliar applications	
The examination papers that were studied offered a wide range of source material that enabled candidates to deploy their knowledge and skills. There were occasional questions that team members criticised but the general view was that the quality of the papers was high. The scrutiny of some questions triggered a discussion of the extent to which it was possible or legitimate to set questions that used wholly unfamiliar settings within which candidates could deploy their knowledge and skills. There was one particular question which sought to do this but where the choice of setting was judged to be too far from the candidates' experience, thus inhibiting good responses by presenting an unnecessary barrier. A little more explanation or the choice of a slightly more familiar setting would have greatly improved the question. However, it was felt that moving candidates into unfamiliar territory was desirable, as it forces them to apply their geographical knowledge and skills in a different context to that which they have learnt, thus getting away from the regurgitation of previously learned case study material with its dependence on recall. Questions that invite candidates to deploy ideas, comparisons or information from cases that they have studied, but applying it to an unfamiliar setting, are likely to be more satisfactory. If such questions then require candidates to synthesise and evaluate information that has come from several sources then they provide a more satisfactory assessment of AO3 and AO4. The team recognised that there should be a limit on the settings that should be used (they must not stray too far from familiar contexts) and that it remains important to assess knowledge and understanding.	**Issues and recommend-ations 2.1 and 2.2**

Box G8: Inclines of difficulty	
Because of the arrangement according to topic area there was no general incline of difficulty through question papers. But there were clear inclines through questions, which frequently had many parts. It was reported that such inclines did exist through problem solving and decisions making papers though the team did not see these papers. Inclines were thought to be essential and that it was important to ease candidates into a question both by the way in which it was presented (see below) and by the nature of its early parts. Inclines were, however, not simple linear structures since it was often necessary to, for example, have a	**Issues and recommend-ations 2.3**

number of basic questions whose information fed into a later part. At all stages it was important that candidates should not require a correct answer to one question part in order to be able to answer the next although marking schemes could adopt the carry-over methods used in mathematics and the sciences. There was some discussion about what made an easy opening question, particularly since the 'define' type of question was reported to cause considerable difficulty for a number of candidates. It is often answered badly because candidates give explanations or examples instead of a definition. The wording of opening questions should be particularly carefully chosen and the subject matter be chosen so as to allow weaker candidates a good opportunity to demonstrate some achievement and to have access to later questions.	51

Box G9: Predictability

The predictability of question papers was considered in the context of several discussions in the course of the two days. It was accepted that there was a degree of predictability in all of the papers seen but that much of this stemmed from their arrangement by topic rather than from the choice of questions used in the topic. The invitation to candidates to deploy information from cases that they had studied was a predictable feature. However, in the best examples seen, a variety of different questions were asked about these case studies, reducing predictability. The best examples of extended answer questions cut down the predictability considerably. It may be the reviser's task to monitor examination papers for predictability. Both formal and less formal methods for checking appeared to be used and are thought to be important. There are expectations of coverage that have to be fulfilled and it is unlikely that any action by awarding bodies will stop teachers from attempting to predict what will appear in an examination. Geography has, in any case, relatively broad specifications that allow examiners reasonable flexibility and reduce predictability. Reducing predictability through format changes is a risky business since candidates may be dismayed at seeing an unfamiliar layout. Unexpected or unfamiliar source material can offset some predictability but there needs to be care in its choice (see above). However, there is a very wide range of good material now available to examiners and its use in new ways might reduce predictability whilst improving the quality of assessment. Thus, an OS map might be used to test the usual range of skills whilst also making a contribution as a source for another question on, for example, glacial landscapes.	**Issues and recommend-ations 1.7, 1.10 and 2.4**

Written paper presentation

Box G10: Answer booklets	
One awarding body used an answer booklet for one of its higher tier papers. All other papers required the candidate to answer in the question booklet and provided spaces for these answers. Team members felt that the use of an answer booklet introduced an additional skill or group of skills associated with the need for a candidate to decide upon the scope and presentation of answers, so making this paper more demanding than those from other awarding bodies. It was not immediately clear why this paper was different from all the others supplied for the exercise. It was thought to be relatively unusual in the current GCSE provision.	**Issues and recommend-ations 3.1**

Box G11: Text and graphics layouts on papers	
There are several issues concerned with the presentation of text and graphics on the examination papers. An emphasis on a word or phrase through the use of a bold or italic font was sometimes generally over-used or used inconsistently within a single paper. Team members felt that there were instances where the wrong word was emphasised. Poor or over-used emphasis detracts from the value of highlighting key words in a way that will assist candidates. Generally, team members did not favour the use of a different font as a way of emphasising text though there were some cases (particularly within graphics) where this was helpful. Team members regretted that some fonts were not as large and they would like but recognised that the examination papers could not afford to be much longer than they already are. However, it is possible that fonts could be enlarged by 1 or 2 points without having much effect on paper length though the use of fonts was normally a matter of awarding body rather than subject practice. Expert advice or research might be valuable in making the selection of the most readable font for this type of application. There were said to be occasional difficulties in reproducing graphics on the coloured paper normally used to distinguish Foundation tier from Higher tier papers. Because this study used photocopied papers the impact of this was difficult to determine. It was essential that graphical and textual source materials should be clearly distinguished from questions. Most of the papers scrutinised achieved this by placing source material in a box; where this was not done there occasionally some	**Issues and recommend-ations 3.2 to 3.5**

difficulties in working through the paper with the possibility that candidates might overlook important information. Where e-marking was used it was essential that candidates wrote answers within the spaces provided. This puts even greater pressure on examiners and awarding body staff to ensure that the answer spaces are well designed. Some answer spaces seemed out of keeping with the question asked; they were sometimes of the wrong length and it seemed usual to provide lined spaces even when only a one-word or numeric answer was needed - it was not clear why such short answer spaces were not more often boxed. Answer spaces do provide candidates with one type of clue about the required answer – an issue discussed more below in connection with candidates' appreciation of the marks available for a question.	

Box G12: Language

There are several observations about the language used in the question papers. The team noted that one method commonly used to distinguish the tiers was the simplification of some language in a Foundation tier paper, when compared with questions (some of which had the same geographic content) at the corresponding Higher tier. The team was not sure whether this was justifiable: weaker performance in geography might not be associated with weaker language performance. Changes in language when geographic content was the same would make it more difficult to compare performances, particularly when attempting alignment at Grade C. There were a few instances where the language was thought to more difficult than it need have been. There seems no reason why language should not generally be as simple as possible and the same for both tiers. The team was unhappy that examiners sometimes justified the use of some terminology on the grounds that 'it is in the specification'; specifications are formal documents written primarily for adult use whereas the language of an examination paper should be close to the language in which the subject is learned – the language of the geography classroom. Inaccessibility of language is then less likely to disadvantage some candidates. However, it is important to distinguish technical terms from common language. Competence in geography must include competence in the use of some parts of the language of geography and this should not be simplified. It would make it much easier for examiners to use suitable language if there were an agreed glossary of essential geographic terms that candidates are expected to know and use. Some specifications either include such glossaries or refer to published glossaries. A particular problem with the language of the examination papers is the use of command words. On some papers these	**Issues and recommend-ations 3.6 to 3.8**

were extremely repetitive and team members felt (partly from comments in examiners' reports) that some of the subtleties in examiners' uses of command words were not understood by all candidates. When used repeatedly they inhibited good language use, often producing stereotyped instructions and question stems that could often have been presented in a simpler and more accessible way. This bears on the wider issue of question presentation, discussed below. Finally, questions actually appear in the form of instructions to questions (that is, ending with "?"). This is clearly a common practice, presumably as a result of a common decision. It does lead, in some instances, to unnecessarily complex or abrupt presentations.	54

Box G13: Question presentation	
Questions in these papers are complex, spread over several pages with a large number of parts. In most cases the presentation of the questions appeared to work well, avoiding the well-know pitfalls associated with complex and lengthy question paper layout. However, there were some areas where team members felt that improvement was possible. There were questions that were introduced in a quite clumsy way, partly (or largely) because of the need to use command words or standard phrases. Thus, for example, a question may begin with the stem "Study Figure x below"; the figure follows and has a title underneath it. Team members felt that candidates' access to the question would be easier if these phrases were linked, such as "Figure x below shows xxxx". Thereafter, some questions would have benefited from linking phrases (or stems) that enabled candidates to make a smooth transition from one section of a question to the next. In some instances phrases of this kind could be used either to make links between apparently unconnected question parts or to help making it clear what type of answer was required. Whilst it was probably true that some candidates didn't read such linking material and just dived into the questions, its value to others justified its use. A few examples were found of questions where the presentation seemed positively inhibiting. Examples are questions where the task to be performed was presented before the source upon which the task was based; questions where items of source material were presented without preamble and questions where no attempt was made to establish linkages from one question part to another. At the same time, the team recognised the difficulties in achieving optimal presentation in all cases and were anxious that improvements in presentation should not result in longer question papers.	**Issues and recommend-ations 3.9 and 3.10**

Marking issues

Box G14: Candidates' awareness of marking criteria	
Command words have already been discussed in connection with the language and text of the question paper. However, these words are traditionally seen to provide information about the type of answer that is required, so that the use of a word like 'define' points to the need for a different type of response than one that would follow from the use of a word like 'explain' or 'describe'. Teachers have had the task of preparing candidates to respond in appropriate ways to these words but it was noticeable that many examiners' reports included comments that suggested that candidates did not produce the type of answer that was indicated by the command word that was used. Of course, in many cases this was probably because they were unable to do so but there is the nagging concern that over-reliance on these words and the subtle shades of meaning that they may convey is lost on some candidates who, given more explicit information, could produce a more creditable response. A general simplification of language, bringing it closer to classroom language use, and the use of more explicit instructions that identify what is required from an answer could reduce the dependence on this stock of command words. Stems that make linkages between question parts and careful management of the answer spaces, both already discussed, may help. The whole should be explicit and not dependent on assumptions about the way in which teachers have prepared candidates for the examination. Clarification of this type should be seen in the context of what is being assessed. It is not being suggested that questions are broken down into simplified bits when it is the candidate's capacity to assemble and present an argument or to evaluate some information that is being assessed. However, the wording of such questions can often point the candidate more clearly in the direction of the answer that is required without such a heavy dependence on stock command words. Candidates already know how many marks are available for a question part, and this provides a further clue about the answer required. The team members felt that marking schemes (particularly when levels were being used) occasionally failed to relate the rewards to the scope and depth of answers as described in the question being asked. It is thought doubtful whether it helps candidates to know which questions are level marked though many of them are probably aware of the existence of levels marking and that it is likely to be used for longer questions.	**Issues and recommend-ations 4.1 and 4.2**

Box G15: Use of level marking	
Team members were concerned that levels marking was sometimes used for questions attracting only 2 or 3 marks. It was thought that levels marking should only be used when questions attract 5 or more marks, perhaps occasionally 4. Markers need to be provided with clear triggers that enable them to distinguish levels before making a judgment within a level. A trigger will often be in terms of the type of skill being deployed in the answer. It would be good practice to include small examples of level marked work in mark schemes, but ensuring that markers understand that these are not the only acceptable answers. The status of levels appeared differ amongst awarding bodies. In some, all three levels were used in both tiers whereas elsewhere three levels were used for the Higher tier and two for the Foundation tier though at least one awarding body was not rigid about this. In the latter case there appeared to be some notional relationship between grades and levels: level 3 might correspond to performances typical of grades B, A and A*. This might not matter unless, in lieu of making the levels scheme explicit, markers were briefed in terms that described a level by reference to what was expected of a candidate with a typical grade performance. That might fail to reward answers that were atypically good or poor. However, the team members recognised that awarding bodies may provide general briefing that was not available to the scrutiny and that might clarify this issue. Without access to fuller information it was not being suggested that all awarding bodies should use levels in exactly the same way (they can simply be regarded as units of account within the marking process).	**Issues and recommend-ations 4.3 to 4.5**

Box G16: Grade C	
It is common for question papers to include some anchor questions that are common to both tiers; these both reflect the fact that the tiered papers are assessing on the same specification and that it is helpful to have such questions when considering the relationship between the papers, particularly, the overlap at Grade C. Some awarding bodies had up to 40% of the questions in common; others had less, perhaps on the grounds that this detracted from the value of having tiers. Elsewhere, questions on the tiers were comparable, though with detailed differences. Equating at Grade C seems likely to be easiest if the anchor questions are identical on both tiers and the marks awarded are the same. Here it will be possible to introduce statistical evidence of performance that can fairly easily be interpreted	**Issues and recommend-ations 3.6, 4.5 and 4.6**

alongside expert judgments. The equating probably becomes a little more awkward (and a little more dependent on judgment) if the marks awarded are different but on parallel linear scales and/or the Foundation tier question is the same as the Higher tier question but with, say, one part removed. Equating seems to get significantly more difficult if, say, the questions have identical content but different wording and most difficult if there are very limited or no anchor questions. In these cases any statistical evidence will be very difficult to interpret. It is difficult to see why question wording should be different on both tiers if it the same content and skills being assessed. Wording that is suitable for Foundation tier candidates should also be used for Higher tier candidates. In general, a higher proportion of common questions seems more appropriate than a lower proportion and, except where it is appropriate to remove some question parts for Foundation tier candidates, the questions should be identical with identical marking schemes. This does suggest that there may be occasions where level 3 may be used for Foundation tier.	57

Appendix 2

History Scrutiny Exercise

This report identifies improvements that may be made to GCSE assessment within History. It is based on a review of the principle findings of the History scrutiny panel. It considers the following issues:

- Specifications in relation to the GCSE criteria
- Overview of written papers
- Detailed review of written papers
- Question topics, marking schemes, source/stimulus materials

Follow a summary of the key issues and recommendations for improvement, and a discussion of the impacts of these, evidence from the History scrutiny panel report is presented in a series of boxes numbered H1-H12. Links between the key issues/recommendations and the more detailed evidence contained in the boxes are clearly identified.

The History scrutiny panel meeting was conducted on 30-31 August 2007.

The scrutiny team for History consisted of:
Glenn Timms (AQA)
John Wright (Edexcel)
Colin Shepherd (OCR)
Carole Mallia (CDELL)
Harry Tolley (CDELL)

Scrutiny outcomes

Issues and recommendations for improvement

1. Specifications in relation to the GCSE criteria

1.1 Specifications along with the related examination papers define the 'versions' of History, which are used for the purposes of the GCSE. This is usually a matter of intense public interest, making changes to History specifications more problematic than with other subjects. Changes resulting from this review will need to be clearly articulated and if needs be stoutly defended. *[See Box H1 below]*

1.2 The GCSE criteria allow the Awarding Bodies to select from a wide range of historical content, though this does not necessarily lead to coherent specifications in which there is a balance between different elements such as British and non-British History, and 'outline studies' and 'studies in depth'. *[See Box H1 below]*

1.3 The specifications (combined with predictable question papers) mean that subject content of can be divided up into small

components (or 'chunks'). These can then be taught as content to be memorised for reproduction in the examination, thus encouraging 'surface' as opposed to 'deep' learning, which is driven by extrinsic rather than intrinsic motivation. *[See Box H1 below]*

1.4   A requirement in specifications to undertake 'outline studies' of periods of about a century would help to ensure that themes, change and continuity were all properly covered in the written papers, and hence addressed more thoroughly in the classroom. Such studies would complement the students' experience of 'studies in depth' of shorter periods of history. *[See Box H1 below]*

1.5   It was suggested that there was an overemphasis in the specifications on testing of knowledge recall, and that there are too few opportunities for candidates to use historical source materials as the basis for 'doing some history', as well as constructing their own narratives based on those investigations i.e. 'writing some history'. *[See Box H2 below]*

1.6   It was suggested that the emphasis in the specifications on testing the ability to 'recall, select, organise and deploy knowledge' was likely to favour some students at the expense of others, notably those with learning styles better suited to the critical analysis of historical sources. *[See Box H2 below]*

1.7   It was recognised that some of the options available in the specifications were potentially controversial and could be problematic to teach in some classroom situations. However, this was thought to be a good reason for their inclusion in GCSE History rather than their exclusion because of the opportunity it gives to schools to further their students' understanding of the modern world. *[See Box H2 below]*

1.8   The limited content of many of the options offers little scope for varying the questions from one examination to another so that they become predictable in both in terms of topics and formats. Teachers who are quick to criticise questions that depart from past papers see this predictability as a virtue. *[See Box H2 below]*

1.9   It was suggested that there is an absence of opportunities in the specifications for candidates at the top of the target grade range (A/A*) to produce developed arguments, solve problems and make judgments based on critical analysis. *[See Box H3 below]*

1.10 There some concern about parity in the levels of difficulty between the papers for different options, the differences often being related to source materials, which can make some choices appear to be more challenging than others. *[See Box H3 below]*

2. Overview of written papers

2.1 Some reservations were expressed about the way in which A02 ('use of historical sources') is currently assessed, the problem being seen as related to the need to assess these 'in their context' as well as critically - analysis, evaluation and interpretation of sources suffering at the expense of recall about their context. *[See Box H4 below]*

2.2 It was suggested that the issues identified in 2.1 might be addressed by means of a source investigation paper bearing some similarities with the decision-making exercises (DMEs) in some GCSE Geography specifications. It was thought that this would not only improve the assessment of the relevant skills but would have a beneficial impact on the curriculum – though there were some doubts as to whether or not the sources should be tied to a particular historical context. *[See Box H4 below]*

2.3 The balance of skills and the distribution of content across the written papers conformed to the requirements of the specifications – though it was noted that some gaps in content coverage inevitably occur from one examination paper to another. *[See Box H4 below]*

2.4 There was a concern that the allocation of some candidates to lower tier papers would deny them access to the types of question, which would enable them to demonstrate latent high-level skills. *[See Box H4 below]*

2.5 With common papers, it was suggested that the structured nature of the questions, whilst providing 'scaffolding' for candidates at the lower end of the target grade range, served to prevent those at the top end from demonstrating achievement at the highest level. *[See Box H5 below]*

2.6 It was thought that the problem identified in 2.5 could be addressed by re-allocating marks in order to better reward higher levels of performance, and by setting more searching sub-sections to structured questions (e.g. ones inviting developed reasoning and argument). It would also counteract the tendency for candidates being able to achieve a high grade by the mechanistic accumulation of marks. *[See Box H5 below]*

2.7 Examples were found in the written papers where superfluous stimulus material could have been removed – not least because it appeared to serve no useful purpose other than to improve the visual appearance of the papers. It was also confusing some candidates who were attempting to use it as source material. *[See Box H5 below]*

2.8 It was noted that candidates have a lot of material to sort through in some of the papers offering option choices, putting pressure on

them to make the right choice. Separate question papers for each option would have the effect of reducing the amount of material candidates have cope with – and lead to smaller general rubrics. *[See Box H5 below]*

3.   Detailed review of the written papers

3.1   As noted in 2.8, in some papers the number of options in the papers of some of the specifications studied creates a large number of questions, which in turn leads to very long papers for candidates to sift through under examination conditions. Providing separate question papers (or booklets) for each option was considered to be preferable. *[See Box H6 below]*

3.2   In terms of their layout, some question papers were seen as being badly designed, with rubrics, source materials, questions and spaces for answers being so arranged that the candidates have to turn over a page in order to link sources to questions. *[See Box H6 below]*

3.3   The practice of printing source materials on a pull out double-page spread (or on fold out pages) so that it can be opened up for use in conjunction with the questions was commended as an effective way of dealing with the problem identified in 3.2. *[See Box H6 below]*

3.4   Instructions to candidates set out in the general rubrics (especially those relating to papers offering option choices under one cover) were generally seen as being too complicated for candidates at the lower end of the target grade range. *[See Box H6 below]*

3.5   By comparison, localised rubrics were seen as being clear and straightforward. However, it was suggested that when addressing issues concerned with the design and presentation of the written papers, thought should be given to the positioning of rubrics as well as to the choice of fonts, and to the use highlighting. *[See Box H6 below]*

3.6   Paper presentation was reasonably clear without being attractive, and the need to turn pages to progress from one section of a question to another (or link source material with the question) was mostly appropriate. However, instances were noted where candidates had to turn the page repeatedly in order to answer questions on sources located on different pages. *[See Box H6 below]*

3.7   In general, the choice of question formats was found to be 'fit for purpose', with structured questions being seen as a suitable format for use in common written papers for all students in the target grade range. However, as elsewhere it was suggested that whilst being helpful to candidates at the lower end of grade range, they might not be sufficiently challenging for those at the very top of that range. *[See Box H7 below]*

3.8 Perhaps because the team were all seasoned examiners, and as such accustomed to seeing question papers at all stages in their development and use, they found the papers to be both predictable and unexciting. *[See Box H7 below]*

3.9 Language levels in all the papers were thought to be appropriate to candidates in the target grade range in terms of sentence clarity and length, and in the way in which ambiguities and complexities in the use of command words were avoided. However, overall little variety or interest was evident in the use of language. *[See Box H8 below]*

3.10 No examples were identified of biases in language usage that would have disadvantaged certain groups – though a few examples were noted where the use of a simpler question format would have helped the setters to avoid using long and complicated sentences (e.g. by separating instructions relating to the use of sources, from the question itself). . *[See Box H8 below]*

3.11 There was scope for improvement across all of the papers in the use of text fonts (size, type), highlighting (bold, underlining, italics) and layout (paragraphing and positioning). . *[See Box H8 below]*

4. Question topics, marking schemes, source/stimulus Materials

4.1 Question topics were found to be predictable - but a fair reflection of the specifications from which they had been derived. However, it was suggested that candidates would not find them to be interesting or exciting. To counteract this there was a plea for more questions about the experiences of people who lived in the historical periods studied, and the inclusion of more interesting source materials offering real issues to be interrogated. *[See Box H9 below]*

4.2 Despite the reservations expressed about the predictability of question topics, team members were critical of instances of what they regarded as questions on 'obscure topics', which would result in candidates being marked for what they didn't know as opposed to what they did know. *[See Box H9 below]*

4.3 The use of a limited range of predictable question topics was thought to be having a damaging effect on the students' experience by encouraging an emphasis on teaching small units of content and the coaching of examination techniques resulting in superficial coverage of the subject, and a lack of depth and coherence in the students' knowledge and understanding. *[See Box H9 below]*

4.4 Marking schemes were found to be: clearly presented, supportive of the marking process and allowed the candidates' achievements to

be appropriately rewarded. The use of levels mark schemes was thought to be better suited to the marking of questions eliciting longer more developed answers than to the short responses asked for in the early parts of structured questions. *[See Box H10 below]*

4.5 Whilst in some marking schemes the guidance lacked sufficient examples, others were so full that they looked more like 'notes for answers' – perhaps an attempt to provide guidance for markers who lack confidence in their own knowledge of particular options. *[See Box H10 below]*

4.6 Reservations were expressed about the use of marking schemes outside of the contexts for which they were devised, especially coaching candidates for written examinations. *[See Box H10 below]*

4.7 Whilst acknowledging that the graphics would have looked better in their original form - as opposed to photocopies – the team was of the opinion that they could be improved. To that end, it was suggested that support materials should be made to look more 'authentic' through the use of colour where appropriate, and by means of high quality 'art work' and reprographics. *[See Box H11 below]*

4.8 It was suggested that supporting materials should be structured into the questions set, and that greater care should be taken in designing the layout of individual pages to ensure that the graphics are positioned in the most appropriate locations in relation to the accompanying text. In addition 'house rules' need to be established for where best to place information relating to the provenance of source materials. *[See Boxes H11 and H12 below]*

4.9 It was suggested that there was a need for greater clarity in how support materials included in written papers are intended to be used by the candidates. It was thought that this could be achieved by means of explicit references to all such items in the wording of the questions, and by signposting where they should be used by their positioning in relation to the text. *[See Box H12 below]*

4.10 It was noted that the language embedded in historic source materials could differ significantly from the current usage with which students are familiar. The problem is particularly evident in cartoons where the embedded language is integral to what cartoonists are seeking to communicate about persons, situations or events. Interpreting sources of this kind may be more difficult for some candidates than others depending on their socio-cultural backgrounds. This was not seen as reason for excluding such items from GCSE History written papers because they are not only a valuable source, but their presence is a reminder of the importance of furthering cross-cultural understanding and the ability to recognise alternative perspectives. *[See Box H12 below]*

Reviewing impacts

5. Impacts on teaching and learning

5.1 It is unlikely that any one of the proposed changes will have a sudden and dramatic impact on the teaching and learning of GCSE History, but cumulatively and over time they should bring about significant improvements.

5.2 The students' learning experiences would be improved by: a reduction in the emphasis in the assessment system on the ability to recall knowledge about small pieces of History; the opportunity to complement studies in depth with an overview of a longer historical period; and, the development of investigative skills and the ability to interpret historical sources.

5.3 The creation of opportunities for students to do some history through the interrogation of 'authentic' sources, and to write some narratives of their own based on those investigations, would serve to encourage 'deep learning' and the development of intrinsic motivation.

6. Impact on the management and administration of examinations

6.1 The changes suggested here would not significantly increase the length of examination papers nor result in an increased volume of material to be accommodated on typical examination room desks. Indeed, the production of separate questions papers for different options in some specifications would serve to reduce the amount of material presented to candidates.

7. Impact on the cost of the examinations

7.1 None of the present proposals are likely to increase examination costs significantly, though question setters might need more time in preparing papers and mark schemes. Some marginal additional costs might be incurred in the initial stages in improving the design and presentation of the written papers. However, they could also help to reduce some costs e.g. less paper would be used in having a separate booklet for each option rather than booklets containing <u>all</u> of options to be sent to <u>all</u> of the centres.

8. Impact on the range of content and skills being assessed

8.1 The proposals here envisage a shift away from the assessment of content knowledge (and the simple reproduction of learned material) towards the assessment of developed reasoning, argument and interpretive skills.

8.2 It is essential that examination papers give as positive an experience as possible to lower achieving candidates whilst ensuring that the achievements of candidates at the top end of the target grade range are recognised and rewarded.

9. Impact on the status and popularity of the subject

9.1 Whilst changes in the written examination papers (and even changes in specifications) are unlikely to have a marked impact in the short term on the status and popularity of History as a GCSE subject it is important that over the longer term they:

    (a)   Enhance (and not detract) from its position.
    (b)   Offer continuity from the History studied at key stage 3.
    (c)   Provide an appropriate foundation for further study post-GCSE.

Evidence

The issues are discussed in more detail in the boxes below in an order that follows that of the scrutiny schedule - though in the discussion itself members of the team frequently linked issues in ways that were more holistic and discussions and cut across the schedule's structure. For each of the issues there is a discussion in the left-hand column and, in the right, a cross-reference to a proposal or recommendation from the list in the section above.

Specifications in relation to the GCSE criteria

Box H1: Ways in which the criteria provide an adequate basis for the specifications	
Specifications are complex statements of intent on the part of the Awarding Bodies that generate them. Within the statutory and regulatory framework they identify the nature of the qualification to be awarded in terms of its content, assessment arrangements and performance requirements. They are public documents addressed to all those concerned with the GCSE in the subject, but in particularly they are directed at and important to teachers (who interpret them for their students) and examiners (who use them as the framework for setting and marking question papers). Alongside the GCSE Code of Practice and the GCSE criteria for History, they define the basis for determining the validity of the grades awarded on the examination. The specifications along with the examination papers therefore, define the 'versions' of History taught, learned, assessed and accredited at GCSE level. This is generally a matter of intense media and public interest, making reform and change, in the opinion of team members,	**Issues and recommend-ations 1.1, 1.2, 1.3 and 1.4**

more problematic than it would appear to be in many other subjects. Consequently, the reasoning behind any proposed changes arising from this review will need to be carefully thought through, clearly articulated and if necessary robustly defended.

The GCSE criteria for History allow the Awarding Bodies to select from a wide range of subject content when devising their specifications. However, this does not necessarily lead to the construction of coherent specifications in which a balance is struck between the different elements such as British and non-British History, and 'outline studies' and 'studies in depth'. With regard to the former, the view was expressed that if British History is so important as to be a requirement, why is it that History is not a compulsory GCSE subject? In respect of the latter it was suggested that 'outline studies' of periods of about 100 years would not only ensure that themes, change and continuity were all properly covered, but would complement and counter-balance the students' experience of 'studies in depth' of shorter periods. Currently, the specifications (and the way that those are translated into predictable examination papers), means that for the purposes of GCSE History, the subject content is divided up into small components (or 'bite-size chunks'). In the experience of the team members, these are then often taught in such a way that the content can be memorised and reproduced in the written examination, thus encouraging 'surface' as opposed to 'deep' leaning driven by extrinsic rather than intrinsic motivation.

It was suggested by the team that the lack of clear definitions of the skills listed in the GCSE criteria for History has resulted in a degree of inconsistency in the ways in which they are explicated in specifications across the Awarding Bodies and subsequently assessed – a problem, which was thought to apply especially to 'interpretation'. This was coupled with the view that there is an apparent overemphasis on testing recalled knowledge at the expense of other objectives. It was also noted that judging from the specifications studied, there are few opportunities for candidates to use historical source materials as the basis for 'doing some history' and constructing their own narratives (i.e. for 'writing some history').

It was thought that as currently formulated, the GCSE criteria for History themselves allow for the development of specifications which meet the needs of all candidates without bias. However, some concerns were expressed about the ability of both tiered and common papers to address the needs of all candidates especially those at either end of the target grade range. In addition, it was suggested that because of differences in individual learning styles, the emphasis on testing the ability to 'recall, select, organise and deploy knowledge' was likely to favour some students at the expense of others (e.g. those better capable of analysing historical

sources critically). Finally, it was recognised that teaching some of the options available in many of the specifications could be problematic in some classroom situations (e.g. 'Changing Relationships: Britain, Northern Ireland and Ireland c1965-c1985' in CCEA History). However, this was judged to be a good reason for their inclusion rather than their exclusion because it gives schools an opportunity to further their students' knowledge and understanding of the world in which they live, and to recognise that there are always alternative perspectives to be considered in relation to the historical events they are studying.	

Box H2: Extent to which the specifications reflect the requirements of the criteria

By and large, the team thought that of the specifications they had examined, the requirements of the criteria had not resulted in any significant constraints or distortions in the assessment as a whole. However, there were some concerns about what was seen as an overemphasis on the ability to recall subject knowledge, and it was suggested that a clearer rationale is needed for the way in which the AOs are distributed in each component. It was also noted that the balance of weighting between the AOs constrained the design of written papers, but did not necessarily distort the overall assessment. Similarly, they were of the view that in general the specifications met the requirements of the GCSE criteria for the subject, and that in turn the relevant written papers were broadly in alignment with those requirements. That said, some concerns were voiced about the extent to which setters (and in turn teachers) were taking full advantage of all the opportunities open to them in the specifications. In saying that they recognised that across the specifications, the subject content in many of the options (especially in so-called 'studies in depth') is so limited that there is little scope for varying the topics on which questions are set from one examination to another. Consequently, it is almost inevitable that after a few years the questions become predictable both in terms of the topics they cover and the formats used by the setters. Such predictability was said by the team to be seen not as a vice but virtue by the majority of teachers who are quick to criticise examiners who take the risk of setting questions that depart in any way from the established 'formula' discernible in past examination papers. Finally, the team considered that for the most part, the overall weighting of content and skills within the written papers was appropriate and in line with the requirements of the relevant specifications. However, some concern was expressed about the variations that can occur across the options in the way in which particular aspects of subject content (such as the 'change over time') are tested.	**Issues and recommend-ations 1.5, 1.6, 1.7 and 1.8**

Box H3: Likely grade outcomes	
By and large, the team members declared themselves to be 'fairly confident' that the specifications they had studied from across the Awarding Bodies would lead to valid and reliable grade outcomes. There were however, a number of reservations. The first related to what was seen as the absence of opportunities for candidates at the top end of the target grade range (i.e. A/A*) to demonstrate their ability to produce developed arguments, solve problems and make judgments based on critical analysis, as opposed to reproducing what they have learned (or 'jumping through hoops without thinking'. The second related to a concern about parity in the levels of difficulty between the different option papers open to candidates. In this respect, it was noted that whilst the question set can be broadly comparable in terms of their difficulty, differences often exist in the related source materials making some options appear at the surface to be more challenging for the candidates than others.	**Issues and recommend-ations 1.9 and 1.10**

Overview of written papers

Box H4: Extent to which the written papers adequately reflect the appropriate parts of the specifications	
The team members made a number of comments concerning the extent to which the written papers they had considered adequately reflected the appropriate parts of the related specifications. In terms of the choice of assessment methods and question types, they found that in general they were appropriate, and as such consistent with the requirements set out in the specifications. However, there were some adverse comments on the ways in which A02 ('use of historical sources') is currently assessed as evidenced by the examination materials under consideration. The problem was seen as being related in no small part to the need to assess the use of such sources not only 'critically', but also 'in their context'. Consequently, the analysis, evaluation and interpretation of source materials can suffer at the expense of recall questions about their context. The team discussed ways of addressing this problem including the possibility of a source investigation paper bearing some similarities with the decision-making exercises (DMEs) found in some GCSE Geography specifications. However, there were some doubts as to whether or not the sources should be tied to a particular context. The feeling was that the ability to interpret sources in their historical context is still an important skill and as such should not be lost. In general the team found that the balance	**Issues and recommend-ations 2.1, 2.2, 2.3 and 2.4**

of skills and the distribution of content across the written papers conformed to the requirements set out in the respective specifications – though it was noted that with regard to the latter some gaps in the detailed coverage will inevitably occur from one year to another.	

The team identified a number of issues with regard to the distribution of content and skills across the target grade range. In particular, they thought that there was a danger that the misallocation of candidates to lower tier papers would deny them access to the types of question, which would enable them to demonstrate their latent high-level skills. In common papers they were concerned that the heavily structured nature of many of the questions, whilst providing a 'scaffolding' for candidates at the lower end of the grade range, served to deny those at the very top end opportunities to achieve at the highest level. It was thought that this problem could be addressed by some re-allocation of the marks in order to reward higher levels of achievement, and by setting more open-ended and searching parts to structured questions which provide opportunities e.g. for developed reasoning and argument.

As far as the design of the written papers was concerned in terms of quantity of material included in them, the team pointed out that there were examples where superfluous stimulus material could have been removed because it served no useful purpose other than to improve the visual appearance of the papers. It was also noted that there was a lot of material 'to be waded through' in some of the papers offering option choices, thus putting unnecessary pressure on the candidates to make the right choice. It was suggested that in such cases it would be better to have separate question papers for each option, which took away the onus of rubric error from the candidates. This would have the effect of not only cutting down the amount of material a candidate had to sift through, but also reducing the length of the instructions (or general rubric) they had to read. Finally, it was suggested that having too many small sub-questions first to read and then to answer can get in the way of reflection and extended reasoning.

Box H5: Content and standard of the written papers

The team members raised several issues concerning the content and overall standard of the written papers. In general they found the demands placed by the written papers on candidates across the target grade range to be appropriate or 'about right' in terms of their levels of difficulty. In particular, with the common papers studied, they found that guidance was given in the framing of the questions, with the aim of preventing the candidates from drifting off the subject. In addition, they commended the way in which relatively simple,	**Issues and recommend-ations 2.5, 2.6, 2.7 and 2.8**

straightforward sections were counter-balanced by more searching parts to the questions. Nevertheless, they reiterated their concerns noted above over the extent to which some of the written papers they had examined allowed the candidates at both ends of the target grade range to fully demonstrate their knowledge and skills. In relation to the extent to which the written papers they had reviewed were likely to lead to valid GCSE grades for all candidates in the target grade range, the team identified a number of issues closely related to matters they had voiced elsewhere. In particular they were concerned over the way in which candidates do not necessarily have to demonstrate high-level knowledge, understanding and skills but can achieve high grades by the mechanistic accumulation of marks. Nevertheless, they suggested that the papers should differentiate reasonably well and with the related marking schemes should facilitate reliable assessment across target grade range.	

Detailed review of the written papers

Box H6: General paper layout	
The team raised a number of issues concerning the overall design of the written papers. In particular, it was thought that the number of options in the papers some of the specifications studied creates a large number of questions, which in turn leads to very long papers. The upshot of this is that there are too many pages for the candidates to sift through under the pressure of examination conditions in order to find the section of the paper, which contains the questions relating to their chosen option. The alternative approach adopted by some of the Awarding Bodies of providing separate question papers (or booklets) for each option was considered to be preferable. A second concern, relates to the layout of the papers. In general, they were seen as being badly designed with rubrics, source materials, questions and spaces for answers often being squashed together, or arranged in such a way that the candidates have to turn over a page in order to link sources to the related questions. The practice of printing the source materials on a separate pull out double-page spread (or fold-out pages) so that it can be opened up and easily studied in conjunction with the relevant questions was commended as an effective way of dealing with this problem. The concerns of the team with the number of option choices candidates were faced with in some of the papers they were asked to study were reflected in the issues they raised concerning general rubrics. By and large they thought that the instructions to candidates set out in the general rubrics located on the front of question papers were too complicated - or 'challenging' for those at the lower end of the target grade range. By comparison, localised rubrics were generally seen as	**Issues and recommend-ations 3.1, 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6**

being clear and straightforward. However, it was thought that when addressing issues concerned with the design and presentation of the written papers careful thought should be given to the positioning of such rubrics, the choice of fonts and the use highlighting - as well as the wording of the information they contain - in order to achieve the maximum clarity for all candidates across the target grade range. A number of issues were also identified with regard to the presentation of the pages within the written papers. In general, they were found to be reasonably clear without being visually attractive (though the individual pages on one of the papers examined were described as being like 'a maze'). Similarly, the need to turn pages in order to progress from one section of a question to another or to link source material with the related question was generally found to be appropriate. However, instances were identified where the candidates would need to turn the pages repeatedly in order to construct answers to questions based on the source material to which they had been directed – a problem, which could have been obviated in one instance by simply changing the order of the question set.	72

Box H7: Choice of question formats	
The team made a number of observations about the choice of question formats in the written papers they examined. In general they found them to be 'fit for purpose' i.e. appropriate to the content and skill being assessed. In particular, they thought that in common papers, the use of structured questions with careful progression through appropriate inclines of difficulty was a suitable format for all students in the target grade range. However, there was some concern that this approach, whilst being helpful to candidates at the lower end of that target grade range, might not be sufficiently challenging for those at the very top. This was coupled with the view that the form of response invited by the questions set, was appropriate for testing recall and description (for allowing the candidates to 'regurgitate what they have been taught') – but less so for demonstrating higher order skills, which need more thought-provoking questions. Perhaps because the team are all seasoned examiners, and as such are all accustomed to seeing question papers at all stages in their development and use, they found the papers to be both predictable and unexciting.	**Issues and recommend-ations 3.7 and 3.8**

Box H8: Language and text	
Several observations were made about the language used in the question papers – though none of them identifying what were regarded as major causes of concern. This means that in general, the language levels used in the written papers across all of the Awarding Bodies were thought to be appropriate to candidates in the target grade range. Sentence clarity and length were both judged to be acceptable and ambiguities and complexities in the use of command words were avoided. However, the need to use easily understandable plain English, and command/instruction words with which the candidates were familiar meant that little variety or interest was detected in the use of language. Similarly, no evidence of biases in usage that would have disadvantaged certain groups was identified. However, a few examples were noted where the use of a simpler question format would have helped the setters to avoid using long and complicated sentences e.g. by separating instructions relating to the use of sources from the question. Finally, with regard to text, it was thought that there was scope for improvement across all of the written papers in the use of fonts (size, type), highlighting (bold, underlining, italics) and text layout (paragraphing and positioning) – matters closely related to issues discussed under other headings.	**Issues and recommend-ations 3.9, 3.10 and 3.11**

Question topics, marking schemes, source/stimulus materials

Box H9: Question topics	
In general, team members found the question topics to be very predictable, but as such a fair reflection of the specifications from which they were derived. In this respect, the question topics chosen were judged to lend themselves to the assessment of the specified content and skills. However, team members were of the view that as a result of their familiarity with the question topics, the candidates would not find them to be particularly interesting or exciting. To counteract this tendency there was a plea for more questions about the experiences of people who lived in the historical periods being studied, coupled with the inclusion of more interesting source materials offering real issues to investigate. Despite the reservations expressed about the predictability of question topics team members were critical of the occasional instances of what they regarded as questions on 'obscure topics', which they felt would result in candidates being marked for what they did not know rather than for what they	**Issues and recommend-ations 4.1, 4.2 and 4.3**

did. With regard to the impact on teaching and learning it was suggested that there was a danger that the limited range of question topics and their predictability has a damaging effect on the students' experience. In particular it was suggested that it could lead to an emphasis on teaching small sections of content and the coaching of examination techniques leading to superficial coverage of the subject, and a lack of coherence and depth in the students' knowledge, understanding and skills.	74

Box H10: Marking schemes	
By and large, the team members found the marking schemes they examined to be clearly presented, supportive of the marking process and allowing the knowledge, understanding and skills demonstrated by the candidates to be appropriately rewarded. That said some observations were made about 'levels mark schemes'; an approach, which seemed to the team to be more appropriate for the marking of questions demanding longer more developed answers than to the short responses required in the early parts of structured questions carrying few marks. Whilst commending the appropriate use of this method of marking, the team noted that some of the descriptors were 'too rambling' making it difficult in their opinion, to identify distinct levels of performance. It was pointed out some of the marking schemes the guidance looked 'too thin' and was lacking in examples. This was counteracted in other cases where it was said that the marking schemes are beginning to look like 'notes for answers' – an attempt, it was suggested to provide guidance for assistant examiners who are themselves uncertain about knowledge and understanding of particular content options. Finally, the team voiced some reservations about the use of marking schemes outside of the examination context for which they were devised, and for purposes other than those for which they were originally intended i.e. to facilitate consistent and reliable marking. As such they are working documents for use in standardising meetings (at which the principles underpinning them, and their practical applications are clarified by senior examiners) and in the subsequent marking of scripts. In particular, there was concern over their widespread use for coaching candidates for the written examinations, and the way they deflect interest away from consideration of wider curriculum and assessment issues at INSET meetings.	**Issues and recommend-ations 4.4, 4.5 and 4.6**

Box H11: Graphics	
The team members expressed a number of concerns both on the production of the graphics (i.e. the diagrams and other illustrations) included in the written papers, their use within the questions set, and their positioning in relation to the text. In particular they were critical of the general standard of production as represented by the graphics included in the sample of written papers they reviewed. Whilst acknowledging that the graphics would have looked better in their original form (as opposed to their photocopied versions), they were of the opinion that they could and should be improved. To that end it was suggested that every effort should be made to make these materials look as 'authentic' as possible through the use of colour where appropriate, and by means of high quality 'art work' and reprographics. They were also recommended that these materials should be structured into the questions set i.e. they would chosen, presented and used to serve clearly defined purposes. In turn, this would mean that in designing the layout of individual pages great care would be taken in placing the graphics in the most appropriate locations in relation to the accompanying text, and establishing 'house rules' for where best to place the information relating to the provenance of source materials.	**Issues and recommend-ations 4.7, 4.8**

Box H12: Source/stimulus materials	
The team identified several issues relating to the clarity of where support materials should be used, their manageability in the examination room, and the suitability/appropriateness of the language relating to their use. In general they were opposed to the inclusion of so-called 'stimulus material' in written papers where they seem to be 'free-standing', and with no clear indication as to how the candidates should or should not use them in answering the questions. The preference of the team was for greater clarity in how support materials were intended to be used by the candidates e.g. by means of explicit references to such items in the wording of the questions, and by 'signposting' where they should be used e.g. by juxtaposing them in relation to the text. With one exception involving the manipulation of three booklets, the source materials, which were embedded within the written papers examined by the team were judged to be manageable by candidates on the size of desks found in typical examination rooms. There was however, one issue noted by the team concerning the uses of language embedded in source materials. If such materials are 'authentic' then the language embedded within them may well differ from the current usage	**Issues and recommend-ations 4.8, 4.9 and 4.10**

with which students are familiar. In a cartoon, for example, the language used will be an integral part of what its creator is seeking to communicate about a particular person, situation or event. Interpreting that message – seeing what the cartoonist is getting at – may be more difficult for some candidates than others depending on their socio-cultural backgrounds. As with controversial subject content, team members saw this as a good reason for including such source materials within GCSE History written papers (and for that matter using them with students in the classroom) rather than excluding them. Finally, some concerns were voiced about the use of boxed quotations from History textbooks as 'source material'. It was suggested that they lacked interest, and that primary sources were preferable.	76

Appendix 3

French Scrutiny Exercise

This report identifies improvements that may be made to GCSE assessment within French. It is based on a review of the principle findings of the French scrutiny panel. It considers the following issues:

- Specifications in relation to the GCSE criteria
- Paper content
- Paper presentation
- Marking issues

Follow a summary of the key issues and recommendations for improvement, and a discussion of the impacts of these, evidence from the French scrutiny panel report is presented in a series of boxes numbered F1-F12. Links between the key issues/recommendations and the more detailed evidence contained in the boxes are clearly identified.

The French scrutiny panel meeting was conducted on 23-24 August 2007.

The scrutiny team for French consisted of

- Ewen Bird (AQA)
- Steven Crossland (OCR)
- Charlie Henry (CCEA)
- Gareth Roberts (WJEC)
- Roger Murphy (CDELL
- John Hamer (CDELL)

Scrutiny outcomes

Issues and recommendations for improvement

1   **The current examination papers**

1.1   Languages do not easily fit the pattern of other subjects. One of the problems, to quote a member of the scrutiny team, is that 'for languages the medium is the message – whereas for other subjects the medium is already there'. As a consequence, while it is possible to test higher-level concepts or reasoning skills in subjects such as history and geography, languages testing at GCSE level is necessarily far more limited in scope. For the most part it is restricted to testing the ability to communicate within a limited range of vocabulary and structure, and within a linear framework. The great majority of candidates lack the range of language to be able to respond other than at a superficial level to questions that, for example, ask them to express a point of view or arrive at a conclusion on the basis of inference. As a consequence such questions often have merely the appearance of higher level testing.

1.2 There were concerns within the scrutiny team that the lack of flexibility in terms of content and structure might have an undesirable impact; for example, in limiting the capacity of the examination to stretch more able candidates, or in making it too predictable. Closely defined content was felt to be appropriate for candidates likely to achieve grade C or below, but a more open-ended approach would more readily enable higher tier candidates to demonstrate mastery of more complex skills and an ability to manipulate language. *[See Boxes F1 and F9 below]*

1.3 Although the team raised concerns in relation to certain areas, discrete skills testing as it exists at present was generally felt to be well done. Nevertheless, there was a large measure of support for introducing moves towards cross or mixed skills approaches to assessment. Such a move would acknowledge the reality that language and the use of language does not function in discrete components, that, for example, listening and responding are necessarily linked. One member referred to interesting mixed skill tasks which he had recently seen in the German Hauptschule and Mittleren Schulabschluss tests in English.

A more holistic approach to skills assessment would, it was argued:

- accord more with the ways in which languages function and hence would be more authentic;

- enable examiners to set more interesting and challenging tasks;

- have a desirable impact on classroom practice;

- be a better option in the event of the introduction of non-tiered papers. *[See Boxes F2 – F5 below]*

1.4 There is a need to introduce strategies for the testing of speaking skills that are likely to have greater validity and reliability than the current approach. Consideration should be given to:

- improving the teaching and assessment of oral skills;

- the possibility of replacing the short formal test with a criteria based 'controlled coursework' model which, amongst other benefits would help to remove the artificiality of the current approach;

- providing better in-service training opportunities for teachers to develop their skills in conducting and assessing oral examinations. *[See Box F4 below]*

1.5 The weighting currently given to writing that puts it on a par with the other skill areas should be reviewed with a view to reducing it. *[See Box F6 below]*

1.6   The current requirement that no more than 10% of the testing should be done using the mother tongue should be reviewed. *[See Box F10 below]*

1.7   There would be considerable benefits in awarding bodies adopting the practice of pre-testing items. Pre-testing would result in greater accuracy in the targeting of items at particular levels/grades. *[See Boxes F2 – F5 below]*

1.8   Care should be taken to ensure that all items present a context which is appropriate given the age range and experience of the majority of GCSE candidates and which is free from cultural and social bias. There was considerable concern about the fact that GCSE MFL examinations increasingly appear to be becoming quite removed from the lived experiences of the majority of 16 year olds in England, Wales and N Ireland. *[See Box F7 below]*

2   Towards a more radical approach

2.1   The teaching of French, and languages generally within schools, faces something of a crisis. Given this, there was a feeling that while there were some improvements that could usefully be made to the existing examination these were not likely to have much impact upon the declining take-up of MFL at Key Stage 4 (currently 31% in England), nor to help to improve the general quality of teaching and learning. For this to happen something more radical was required. Consideration was given to two areas – the perceived advantages of e-assessment/learning; and the alignment of GCSE testing with existing alternative approaches to the assessment of MFL. In the second of these two areas it was noted that much of the benefit of graded testing movements in MFL (Pennycuick and Murphy, 1988) had been lost in recent years by not adopting the best aspects of this approach within GCSE.

2.2   E-assessment/learning

  - Paper based examinations in MFL look increasingly old-fashioned and have limited capacity to motivate students. E-assessment and learning is particularly appropriate in the teaching of languages.
  - Incorporating the use of e-mail, video-conferencing, DVD, video and other technologies offer exciting opportunities both in the classroom and in the format of examinations.
  - On-line, on-demand assessment would provide opportunities not only for improved and enriched feedback but also would again be likely to aid motivation.
  - Such an approach would more readily support requirements for the use of authentic materials and, were it to be adopted, mixed skills testing (see 1.3 above).

2.3   Relationship with alternative assessments

- Consideration should be given to means of aligning GCSE examinations more closely, in terms of approaches to testing and qualification equivalence, with initiatives such as Asset Languages and the Languages Ladder.

2.4   Various other issues for improving French GCSE examinations were suggested. These included:

- the use and quality of **illustration** on papers, both in the provision of icons for the answering of questions and in the inclusion of "contextualising" illustrations which help provide the subject matter of a given exercise thereby giving useful support to candidates. Variation of illustration to include drawings, cartoons, photographs etc. thereby enhancing the attractiveness of presentation

- variation in the use of **fonts** including the use of bold type and suitable devices to emphasis key points in rubrics.

- careful consideration of the **readability** of rubrics and instructions in both English and French.

- catering for the **less-able candidates**: it is essential that the original intention of GCSE providing a test for candidates from Grade G upwards rather than simply Grade C and above, which is becoming increasingly evident, should be reinstated.

- All **marking schemes** should include a combination of *general* marking guidelines and *specific examples* taken from candidates' work on the current examination. Schemes as given to examiners should be reproduced fully in the report to Centres. In addition, students should always be shown how marks are apportioned within different parts of written examination papers.

References
Pennycuick, D and Murphy R (1988), The Impact of Graded Tests, Falmer Press: Lewes

Evidence

The issues are discussed in more detail in the boxes below in an order that follows that of the scrutiny schedule; however, panel members frequently linked issues in such a way that they are best presented as more holistic discussions that cut across the schedule structure. For each issue there is a discussion in the left-hand column and, in the right, a cross-reference to a proposal or recommendation from the list in the section above.

Specifications in relation to the GCSE criteria

Box F1: The specifications – overview	
The specifications for French, as for other languages, are tightly defined. The National Curriculum for Modern Foreign Languages (MFL) and its supporting documents (eg non-statutory guidance), together with the GCSE criteria for MFL specifies the grammatical and linguistic structures and the vocabulary that candidates have to master, and the various contexts within which they are required to demonstrate their knowledge and understanding. Consequently there is a good deal of similarity between the specifications and the examination papers offered by the different awarding bodies, making them in the words of one member of the team 'virtually inter-changeable'. The four skills are each tested discretely; the content of the topic areas is the same; and the range of test types used is very similar.	**Issues and recommend-ations 1.2**

Box F2: The skills – Listening	
All the passages used in assessing listening skills are chosen by the principal examiner from authentic material, adapted as necessary and presented by native speakers. There is a possible issue about the way in which candidates may need to exercise other skills in order to make sense of and respond to the accompanying questions (e.g. tests of listening comprehension also inevitably test skills of reading and/or writing); but for the most part the team were satisfied that awarding bodies have appropriate strategies in place to deal with this. Generally the assessment of listening skills is well received and, within the present structure, there was felt to be little room for improvement, other than the possibility of reviewing the testing of higher-level skills.	**Issues and recommend-ations 1.3, 1.7**

Box F3: The skills – Reading	
Discussion on the assessment of reading raised a number of points that were similar to those that emerged during consideration of listening. There were instances where questions were set that could be correctly answered without reference to the relevant passage. Clearly this is not good practice.	**Issues and recommend-ations 1.3, 1.7**

Box F4: The skills — Speaking	
This was the skill area where it was felt that there was the greatest need for improvement – in terms both of assessment and classroom practice. The particular points raised were: • Existing requirements place too much emphasis upon role-play. There is abundant evidence that existing requirements for role-play lead to limited and narrow classroom practice. • There is too much regurgitation of prepared responses and routines; and there are a regrettable number of instances of malpractice where candidates are clearly being prompted in ways that are unacceptable. • On occasion candidates are severely disadvantaged because teachers fail to take account of the prescribed hurdles. • The proper and effective conduct of oral examinations as currently specified requires a level of linguistic and assessment skills that not all teachers of French possess – especially when, as may often be the case, they are non-specialists. Although the awarding bodies offer opportunities for training, these are not necessarily taken up. The team recommended, therefore, that consideration be given to alternative approaches to assessing this skill.	**Issues and recommend-ations 1.3, 1.4, 1.7**

Box F5: The skills — Writing	
Some team members were concerned that undue emphasis was placed on writing by all the awarding bodies – although more markedly so by some. The team was, however, divided on this. Alternatively, it was argued that not only does writing play a considerable role in effective learning, but also the development of writing skills at GCSE was an essential part of preparing students to progress to AS/A2 level. Although there have been considerable improvements and candidates produce some impressive pieces of writing, the specifications were generally felt to constrain examiners from setting questions that might encourage more adventurous and imaginative responses – especially from higher tier candidates. The greater flexibility and wider spread of topics that was possible in coursework was better in this respect, although coursework did raise other issues to do with its management and control. Ways should be sought of capturing in the examination paper some of the advantages offered by coursework for generating more exciting writing. There was some support for crediting writing produced by candidates in other components of the examination.	**Issues and recommend-ations 1.3, 1.7**

Box F6: Weighting ranges	
Currently all four skill areas receive an equal weighting of 25%. The regulations are in the process of being amended to give awarding bodies some – albeit limited – discretion on weighting. The scrutiny team felt, however, that this was unlikely to have a significant impact – particularly in relation to the issue about the weighting given to writing.	**Issues and recommend-ations 1.5**

Paper content

Box F7: Question topics	
Where there were concerns about the context in which questions were placed, these related to issues of artificiality and possible cultural/social bias. For example, tasks requiring candidates to write a letter making a reservation in a hotel or an article in French for a school magazine were felt not to be realistic. Equally, questions presented in terms of 'during your visit to France' or which referred to 'your French pen-pal' could be seen as embodying a certain social bias and, indeed, remoteness from the actual lives of today's GCSE candidates who mostly neither write letters or have pen pals!	**Issues and recommend-ations 1.8**

Box F8: Inclines of difficulty	
There was a consensus that the papers scrutinised allowed candidates, at both foundation and higher tiers, to demonstrate their knowledge and skills. For the most part this was achieved via suitably graded questions/tasks across the particular paper following an essentially linear incline of difficulty. Not all papers, however, followed this model. Some adopted a design more akin to a 'peaks and troughs' model, and there was some discussion amongst the team about which was the more appropriate model. With both models, the team recognised the difficulties inherent in ensuring that the increased level of difficulty between one question and the next was not so great as to prove too daunting – particularly in the earlier parts of a paper where it was important to ease candidates into it. They acknowledged that not all examiners or papers got this right all of the time!	

Box F9: Predictability	
Predictability and a degree of repetition are to an extent both inevitable and desirable. The report from one of the awarding bodies, for example, noted that 'the papers contained the familiar range of topics, text styles and test types that are the regular features of this test – no bad thing!' With some justification schools express concern if there are elements within a paper that are perceived to be outside the norm. As with other aspects of assessment, however, there is a balance to be struck. Given the tight parameters within which they are required to operate, examiners have to be extremely skilful if they are to succeed in producing papers that are sufficiently fresh to be interesting and challenging and sufficiently familiar to avoid provoking complaints. In reality the team felt that examiners in French and modern languages generally had less flexibility than examiners in most other subjects, and this could increase the danger of papers being very predictable and possibly, therefore, lessening the validity of them as tests of candidates' knowledge and understanding.	**Issues and recommend-ations 1.2**

Paper presentation

Box F10: Use of mother tongue and target language	
The current requirement is that no more than 10% of the testing should be done using the mother tongue (e.g. English or Welsh). Members of the scrutiny team generally felt that this was too restricting. They highlighted, for example, the difficulties that arise in testing reading and listening comprehension in any depth where the questions and answers were in the target language rather than the candidates' mother tongue. Similarly, they pointed to the problem for examiners of avoiding giving too much of a lead to candidates where the stimuli presented for a piece of writing were in the target language rather than the mother tongue.	**Issues and recommend-ations 1.6**

Box F11 : Question types/format	
There were concerns that the scrutiny had revealed a lack of variety of question types in many of the papers, and there was a general desire for awarding bodies to find ways of remedying this. Currently, for example, there was commonly an undue reliance on certain types of 'tick box' response items (e.g. matching pictures to words) - particularly with foundation tier papers.	

Marking issues

Box F12: Candidates' awareness of marking criteria	
For all the papers the majority of rubrics are required to be in the target language and have been reduced to a minimum. They are similar for each awarding body, are used year-on-year and schools are used to them. Generally the test types used are such that they do not need long explanatory rubrics. For the most part the team were satisfied that candidates were sufficiently well informed about what was required of them. Only in one instance was concern expressed about a rubric that failed to indicate that the quality of language used in the response was significant in the marks allocated to a question.	