

Predictability studies report

A study of GCSE and GCE level examinations



August 2008
Ofqual/08/3866

Contents

Executive summary	1
Background	1
Methodology	1
Findings	2
Introduction.....	4
Background.....	4
Methodology.....	6
Scope of the work	6
Personnel.....	7
Overview of the process.....	7
Materials	7
Phase one – materials review	7
Phase two – centre visits	9
Phase two – desk-based research	10
Phase three – review of candidate work	10
Limitations of the study.....	12
Phase one – materials review	13
Phase two – centre visits	14
Phase two – desk-based research	14
Phase three – review of candidate work	15
Findings.....	16
Study 1: GCSE history	16
Study 2: GCSE media studies.....	17
Study 3: GCSE modern foreign languages	18
Study 4: GCE English literature	19

Study 5: GCE geography	21
Study 6: GCE law	23
Study 7: GCE psychology	24
Study 8: GCE physics	25
Study 9: GCE religious studies	26
Overview	28
Appendix A - Specifications and question papers	30
Appendix B – Contributors.....	36
Appendix C – Consultants’ templates.....	39
Appendix D – Common schedule of questions.....	48
Appendix E – Subject issues	55
Appendix F – Resources and materials for each phase	61

Executive summary

Background

Ofqual carries out regular work investigating examination standards. A comment often made during these comparability studies is that their findings might be affected by the extent to which the examinations reviewed are or are not predictable. The studies are already highly complex and resource intensive, so rather than complicate them further it was decided to carry out a specific piece of work to investigate predictability.

This exercise took the form of a series of linked investigations carried out in 2007. The research was planned to identify issues associated with predictability and formulaic questions within examination papers, and to investigate the consequences of such issues. The work sought to identify the impact of predictability on candidates' experience within the context of GCSE and GCE level examinations.

For the purposes of this exercise, a predictable examination was defined as one in which the nature of the examination paper could be sufficiently accurately predicted to mean either that the examination was not testing the full range of content expected or that it would not be assessing the assessment objectives as defined in the specification. In particular, a highly predictable examination would tend to reward recall of knowledge even where it was ostensibly assessing analysis or evaluation. It was noted that such an examination would be likely to have two particularly unfortunate effects: it may well affect the teaching and learning for the subject, significantly reducing its validity; and, given the context of the work, it would mean that judgements about comparability would in all probability change if the level of predictability were taken into account.

In this context, two important assumptions were made. The first was that predictability is not automatically a bad thing; rather, it is important to distinguish between desirable and undesirable predictability. The second assumption was that any exploration of predictability as defined would need to involve both syllabus materials and candidate responses, and that the work would benefit from hearing the candidates' own views.

The focus of the study was purely on current examinations. Whatever the findings, it is not possible to conclude that the situation is better or worse than it used to be.

Methodology

In selecting the qualifications to be studied, recent monitoring reports were reviewed for cases where predictability had been identified as a possible concern, and Ofqual staff and independent consultants with extensive experience of assessment identified

current examinations where predictability might be an issue. The nine separate studies covered:

- GCSE modern foreign languages (French and German), history and media studies
- GCE psychology, English literature, geography, law, physics and religious studies.

The studies comprised three phases carried out by independent consultants, with the second and third phases dependent on the outcome of the first:

- a review of the examination papers and mark schemes.

If predictability was found to be an issue at this stage, this was followed by:

- visits to schools and colleges to obtain the views of teachers and learners
- a review of selected candidate work from the 2007 examination.

At the end of phase one, the consultants used their findings to make predictions about the summer 2007 question papers. During phase three they looked for evidence to support or counter these findings.

Importantly, the range of techniques used to investigate predictability allowed for any hypotheses to be carefully triangulated. For example, the interviews with teachers and learners were critical in deciding whether or not any perceived problem with the papers was having undesirable consequences in terms of teaching the test rather than the subject.

Findings

In order to understand the nature of the findings properly, it is important to see them in context. The subjects and particular specifications that were included in the work were those which had been identified as being the highest risk in terms of being highly predictable. This part of the exercise had already suggested that concerns in the area were not actually very widespread.

Bearing this in mind, the study suggested that although some issues with examinations being over-predictable were identified, these were relatively few. In particular, the study threw a rather different light on the issue of formulaic questions. In fact, in many circumstances it was judged preferable to use consistent wording where candidates were being expected to display the same skills. Problems were seen where there were small variations in wording across series, especially where these led to some lack of clarity in the use of command words such as 'describe' and 'explain'. To the extent that this use of consistent command words is using a formula,

the study suggested that formulaic questions were desirable. However, as explained below, where other factors were also in play, formulaic questions contributed to the problems.

Where problems with predictability did occur, they tended to stem from interactions of several effects. For example, there were cases where the nature of the examination paper meant that learners needed to prepare significantly less of the content than the specification implied. Here, formulaic questions were an issue, since they further narrowed the scope of what was being assessed. In fact, the main cause of predictability in all cases arose from the way the content was specified. This was often over-precise, reducing the flexibility available to the question setters. It should be noted that, in the cases where there were issues with predictability, the examinations in question either have been revised in ways that significantly reduce the issues, or are in the process of being revised, and the problems are being taken into account in that revision.

However, it was recognised that there was a fine balance to be maintained between being properly explicit in how the subject content was specified and that explicitness leading to predictable question papers. This was in fact the central finding of the work: unpredictable question papers are just as poor in assessment terms as over-predictable ones. Both place too much of a premium on examination preparation and technique rather than being fair tests of the prescribed knowledge, understanding and skills.

Introduction

The following report gives an account of a series of linked investigations carried out in 2007. The research was planned to identify issues associated with predictability and formulaic questions within examination papers, and to investigate the consequences of such issues. The work sought to identify the impact of predictability on candidates' experience within the context of GCE and GCSE level examinations. Although the work focused on specific subjects and, indeed, specific question papers from specific awarding bodies, the main purpose of the project was not to apportion blame for any shortcomings identified but to gain a greater understanding of the kinds of factors that might affect predictability. This report will reflect that main intention by concentrating on what has been learnt about the factors rather than naming names. However, where specific examinations raise significant issues, Ofqual will explore with the relevant awarding bodies any appropriate remedial action.

It was noted that there is probably not an ideal time for such work, but that in this case the work would be timely in many ways. It was, admittedly, too late directly to influence the development and accreditation of revised A level specifications for first teaching in September 2008, but because it was underway during that process it was possible to ensure that those responsible looked closely at the issue. It also provides useful evidence for future monitoring work on the new specifications. In terms of revisions of GCSE for first teaching in 2009, the work is particularly timely, since it has been possible to feed in any important findings well before revised specifications are submitted, and even to review proposed criteria in the light of work in progress.

Background

Ofqual carries out regular work investigating examination standards. This work focuses mainly on comparability of standards between different specifications in the same subject and the maintenance of standards over time. However, it also considers other dimensions of comparability, such as between-subject or between-qualification. A regular comment made during such exercises is that findings might be affected by the extent to which the examinations reviewed are or are not predictable.

However, the exercises are already highly complex and resource intensive, and they are designed as snapshots, even where there is a longitudinal aspect as with standards over time. The lack of knowledge about predictability is recognised as a limitation of the work, but the sheer scale of the task means that no attempt is made to address it. Instead, it was decided to carry out a specific piece of work to investigate predictability.

An initial literature review suggested that although many examination authorities across the world recognise predictability as a possible issue, there seems to have

been very little attempt to investigate the issue. Consequently the work was designed from scratch. A number of assumptions were made from the outset.

Perhaps the most important of these is that predictability is not automatically a bad thing. An examination which was highly unpredictable to candidates would almost certainly place an undue premium on confidence and examination technique rather than levels of relevant knowledge, understanding and skill, and would thus be of questionable validity. This was an idea expressed by and to the participants at regular stages of the work. As a result, it ran through the whole design of the work, including selecting examinations for review where the risk of overly unpredictable questions had been identified.

The key way that undesirable predictability was defined was against two related but distinct features. First, an examination would be too predictable if it meant that the skills being tested were probably different from those implied in the specifications (typically recall would substitute for higher-order skills). Essentially, this would mean that the level and form of predictability would be likely to affect teaching and learning in the subject. The second feature reflects the reason for carrying out the work: if an examination were identified as too predictable, it would mean that conclusions about its comparability with other examinations would be likely to be affected by the findings.

The second assumption, which follows directly from the first, was that any exploration of predictability would need to involve careful consideration of both syllabus materials and candidate responses. An extension of this was that the work would greatly benefit from hearing the candidates' own views (and indeed those of their teachers): much of the work necessarily made use of expert consultants to identify potential sources of predictability. But because, by definition, a candidate's experience of an examination is a single event,¹ it was important to understand the extent to which any features identified would be reflected in the way students prepared for the examinations.

One final detail needs to be understood at this point. Although much of the impulse for the work came from reviews of examination standards over time, the focus of this exercise was purely on current examinations. Whatever the findings, it is not possible to conclude that the situation is better or worse than it used to be.

¹ Except for limited opportunity for resitting.

Methodology

Scope of the work

As already noted, it was recognised from the outset that it was necessary for examinations to be predictable to an extent (the term used to capture this idea in the study was ‘familiar’) and that there were several possible sources of predictability in examinations, often arising from the need to ensure that familiarity. The aim of the work was to consider as many of these sources of predictability as possible.

The process by which the sources were identified was three-pronged. First, Ofqual staff with extensive experience in assessment identified possible features in current examinations where predictability might be an issue. Second, recent reports from all examination monitoring activity were reviewed for cases where predictability had been identified as a possible concern. In addition, Ofqual’s most experienced assessment consultants were asked whether there were issues of predictability in their specific areas of expertise and, if so, where they may best be exemplified in current examinations. The information arising from these sources was then synthesised to produce a list of specific examinations to review. The aim was to ensure that the work overall looked at all the possible sources of predictability identified across both GCSE and A level examinations and across a range of subject areas. Inevitably, this was tempered by resource implications. In all cases, the first stage of the work involved other examinations in the same subject and at the same level to act as controls.

In the end, the scoping for the work led to a review of examinations involving:

- a range of question types (eg short answers/essay style)
- papers with different structures (eg optional question routes/mandatory questions)
- questions with different assessment aims (eg questions aimed at knowledge and understanding, questions addressing skills such as application or evaluation, and questions assessing synoptic understanding).

The research comprised nine separate studies covering the following subjects and levels:

- GCSE level in modern foreign languages (French and German), history and media studies
- GCE level in psychology, English literature, geography, law, physics and religious studies.

More detail about the particular specifications and question papers reviewed is provided at Appendix A.

Personnel

For each study, review teams of two or three consultants were engaged. The consultants had both relevant subject knowledge and experience of the examination system. In each case, one of the team was asked to act as lead. This involved collating responses from the team and producing a summary report at each stage, and advising on subject-specific issues as necessary. A full list of the consultants is provided at Appendix B.

Each study was managed by a member of Ofqual staff, and the whole project was overseen by a project manager and project director, who were also Ofqual staff.

Overview of the process

The specific focus of the work meant that it had a number of unusual design features. These will be explained in more detail below, but it is important to understand the overall design at this stage. Essentially the work was divided into three phases, with the exact nature of phases two and three – indeed, the decision whether or not to extend the work – dependent on the outcomes of phase one. Phase one involved a review of the examination papers and associated mark schemes to gain a full understanding of whether or not over- or under-predictability was really an issue and to decide whether and, if so, how to extend the work. Phase two involved obtaining the views of teachers and learners, as far as possible, by visiting schools, colleges and universities to understand approaches to preparing candidates for the examination. Phase three involved the original panels of reviewers examining selected candidate work from the 2007 examination, to determine whether the features they had identified were having the anticipated effects on candidate performance.

Materials

For phase one, awarding bodies provided the specifications, examiners' reports, examination papers and associated mark schemes from the past four examination series for each of the question papers involved in the review. In addition, to assist with phase two, they also provided some details of schools and colleges entering the relevant examination in summer 2007. For phase three, they provided candidates' scripts from summer 2007, according to a set of requirements specified by Ofqual.

Phase one – materials review

Phase one involved consultants reviewing awarding body specification materials across four series of the examination (eg 2003–6). The process started with a briefing meeting, where Ofqual staff explained the intended approach and associated documentation to the consultants, and all involved explored the issues to ensure a

common understanding of the documentation and agreed the exact scope of the work. In some cases, further materials had to be obtained.² Consultants then worked independently at home making judgements about the predictability issues identified on the specified examination papers. These judgements required consultants to rate the papers against a number of factors using a scale of between one and eleven (very unpredictable/appropriate/very predictable) and record their ratings, together with an explanation for the reasoning behind them, on a form developed for the purpose. A template of the form used in this phase of the study is provided at Appendix C.

Consultants considered:

- the extent to which the structure of the examination paper significantly affected the subject content that had to be studied
- the extent to which the nature of the question papers was over-specified in the specification
- whether the nature of the questions within an examination meant that some skills or knowledge did not have to be developed to the extent indicated in the specification
- whether there were clear patterns to the questions across series which allowed for much more limited coverage of the content than the specification seemed to imply.

Consultants were also asked to consider the possible impact, both positive and negative, of textbooks and other awarding body endorsed support materials. This inevitably depended on such materials being both identified and readily available. The final part of this phase involved the consultants using their findings to make predictions about the summer 2007 question papers.

Consultants then sent their completed forms to the lead consultant who collated them to produce a short summary report. This was followed by a second meeting of all involved in the study, including the project manager and project director. This took place after the summer examination. At the meeting, consultants considered the summary report to ensure that there was real consensus about any issues identified and, in particular, that any apparent areas of disagreement did not arise from slightly different interpretations of the forms. It also allowed the opportunity to review the

² For example, at the initial meeting, consultants sometimes agreed that they needed materials from more than four series to make their judgements; or they identified other materials that were necessary to determine whether or not predictability was an issue.

predictions about the summer 2007 question paper against the reality. In addition, the meeting provided the opportunity to identify the next steps, if any, to collect evidence about the consultants' findings. If no significant issues were identified – that is, the predictability of the questions/question papers was considered appropriate – then work stopped at this stage for that specification and unit.

Phase two – centre visits

Where significant issues of predictability were identified, the next stage was to conduct a series of centre visits. However, such centre visits only took place where it was agreed that the specific issues identified could be supported by evidence gained from interviews with teachers and learners. The interviews were held in the 2007 autumn term.

The aim of the centre visits was to explore the experiences of centres in preparing their students for the summer 2007 examinations and what guidance and/or materials provided by awarding bodies they drew on in deciding that preparation. The interviews were focused on the specific units under review. Teachers and learners were interviewed separately.

The teachers interviewed were those who had taught the particular unit as part of the course for the summer 2007 examination. The learners interviewed were either sixth formers (for GCSE or AS units) or undergraduates (for A2 units) who had taken the summer 2007 examination for that particular unit. Interviews were semi-structured, using a common schedule of questions for each visit. The general framework of the interviews was common across all studies, with some variation for specific issues within particular subjects. Both teachers and pupils were provided with a copy of the summer 2007 examination paper to refer to. The common schedule of questions is provided in Appendix D.

This was a small-scale activity, and the intention was to obtain a clearer idea of how far the understanding of issues identified by experts specifically considering predictability was shared by teachers and, perhaps more importantly, those actually taking the examination. For the visits, the centres with reasonably large entries for the examination in question were identified in terms of size, centre type and performance at GCSE and A level (using the websites www.schoolsfinder.direct.gov.uk and www.dfes.gov.uk/performance/tables/schools_06.shtml). For logistical reasons, centres for GCSE examinations were required to have a sixth form where pupils could continue their studies through AS and A level qualifications, or equivalent (not necessarily in that subject). In addition, some centres were able to arrange for pupils who had left that centre to be interviewed. For some A2 units, teachers within centres were interviewed, but the candidates were interviewed in higher education institutions.

Each teacher and pupil interview was documented in a short report.

Phase two – desk-based research

For a small number of examples, phase one did suggest that there were issues with predictability but it was agreed that centre visits would not usefully inform understanding of those issues. In these cases, consultants undertook desk-based research. The research took the form of carrying out an analysis of the 2007 question papers and mark schemes in terms of their predictability, identifying the most and least predictable items, and estimating, in terms of low, moderate or high, the difficulty of each item taken on its own. This analysis was then compared with item-level data to see if the statistics bore out the hypothesis that predictable questions will be likely to exhibit either a high mean or an unexpectedly low standard deviation, taking into account the difficulty of the question. Where there was question choice, the hypothesis was supplemented by the idea that highly predictable questions would be likely to prove popular.

Consultants therefore compiled a short report on whether there was any evidence in the item-level data to suggest whether:

- candidates were gaining large numbers of marks on certain questions rather than others
- candidates were gaining more marks for certain assessment objectives (eg recall of knowledge)
- candidates were gaining a significant proportion of marks on areas of content that were identified in phase one as predictable
- candidates were gaining higher marks on certain types of question
- candidates' marks were bunched.

Phase three – review of candidate work

From the way that predictability was defined at the outset, it was recognised that it would be necessary to consider candidates' answers to the papers as evidence to support or counter the consultant's findings from phase one. To this end, a script review was held. The aim of the script review was to ascertain whether answers did in fact contain evidence that candidates found specific examinations, or questions within those examinations, predictable or unpredictable.

The exact nature of the scripts to review was discussed at the second meeting. It was agreed that the kind of evidence needed would probably show up within a centre, and as a result awarding bodies were asked to supply the work from all candidates within a selection of centres for the relevant examinations. For each examination

involved in this phase, 15 centres were selected. As with phase two, centres were chosen to cover a range in terms of size, centre type and performance at GCSE and A level, using the government websites identified in phase two. Where it was considered relevant, centres from phase two were included in the sample.

Consultants worked independently on whole centres, making judgements about whether candidate responses showed the kinds of features that might be expected in view of the specific predictability issue(s) identified. Consultants were asked to review scripts across a range of marks within each of the 15 centres chosen, continuing work on a centre until they were confident that the answers either confirmed the expected patterns or that they did not. The number of candidate scripts considered within a centre therefore varied, as consultants stopped reading scripts when they were confident about their judgement for a centre/across all centres on a particular predictability issue. If, after looking at a range of scripts, there was no clear evidence of a problem at least within some centres, predictability was not considered to be a significant issue.

Other materials were also made available to consultants to reference during the script review. Materials included specifications, question papers, mark schemes, examiners' reports, reports from phase two centre visits, awarding body endorsed support materials and, where available, item-level data.

Each consultant independently recorded their judgements on a form designed to ensure that judgements were not distracted by other features of the examination papers, the candidates' answers, or the way they had been marked. The script review concluded with a plenary session whereby consultants shared their findings, came to a consensus view if possible, and agreed examples to illustrate that view. The lead consultant then used the forms together with the outcomes of the plenary to compile a short report summarising the findings. A template of the form used in this phase of the study is provided at Appendix C.

Limitations of the study

The predictability study was carried out between May 2007 and December 2007. This research was different from other projects relating to standards undertaken by Ofqual. As a result, much of the methodology had not been tried and tested in this format and there was little to compare it to. In many ways, therefore, much of what was done was by way of a pilot.

The documentation used throughout the project is a particular aspect of this. Reviewers were asked to record their judgements about predictability on a set of newly developed forms. These forms could not be formally piloted within the timescales available. As a result, there were some minor differences in how the consultants interpreted the forms. However, in all cases differences related to the reviewers' interpretation of forms rather than the identification of, and severity of, predictability issues. Examples of all the templates for the forms used at each phase are provided at Appendix C.

In addition, there are inevitable limitations associated with the scale of the work. For a start, it was selective in the range of examinations studied. It was possible to mitigate this to some extent by planning: a review of recent monitoring and comparability work enabled some real targeting. However, it is clearly the case that using work which had not asked those involved a particular question to help answer that question may result in flawed understanding. It is possible that the study failed to address a particular source of predictability at all. The flexibility of the design of the study throughout, and in particular the opportunity for the reviewers to adjust the scope of each study, will have helped to reduce this risk.

The relatively limited scale of the work, together with the means by which the examinations involved were targeted, has a second important effect on the findings, or rather on how to interpret them. The examinations included in the study were not a random sample, but one specifically selected to gain greater insight into the issues surrounding predictability. As a result, it is easy to extrapolate from the findings to assume that they apply to the whole set of examination papers. Quite the opposite is true: the fact that it was not particularly easy to identify appropriate papers for investigation, taken together with the fact that not all the cases investigated led to serious concerns, is reassuring.

Limitations on resources, time, personnel and materials for each stage of the process were another drawback. Thus the number of reviewers was restricted to three. There is certainly a case for using more, especially as it was not possible to use three in all phases of all studies. In this case, it does not appear to have been a significant drawback. This seems to have been partly because the experts used were being asked to address an unfamiliar issue and found the process interesting and challenging. In addition, there was almost always a strong consensus between the

reviewers about the issues being identified at each stage. It is worth pointing out that the few occasions when there was not such a consensus led to a profitable exploration of the topic and a much clearer understanding of the implications of predictability.

The centre visits were very limited in scale in the design of the work. It is a major logistical problem to arrange visits of this type, depending as they do on the willingness of centres to participate, the ability to bring suitable candidates to the visit, and the availability of dates convenient for all involved. Inevitably, too, the sample of centres visited is partly self-selecting and cannot be assumed to be representative. However, these were never intended to be other than snapshots of practice, enriching understanding of the implications of the findings rather than materially affecting them. In this role, they admirably succeeded.

There was also an inevitable limitation on the range of scripts looked at. This was partly addressed by discussing with the reviewers, once they had a sense of the issues in advance of the meeting, how to select work so that it would be most helpful. It is also worth pointing out that the total volume of candidate work seen in each study was well over 100 scripts for each unit involved. With minor exceptions which will be discussed below, reviewers were confident in their findings from this phase.

For both the centre visits and the script review, there was also a real effort to ensure balance, for example, across centre types or the range of performance seen. Further explanations are provided below about specific aspects of each phase of the study.

Phase one – materials review

In order to maximise the effectiveness of the work, the scoping process involved a review of identified issues to ensure that the study covered as wide a range as possible. Thus, within these studies, the examination papers illustrated a range of question types (eg short answers/essay style), papers with different structures (eg optional questions/mandatory questions) and questions with different assessment aims (eg questions aimed at specific skills, questions testing knowledge and understanding, and questions testing synoptic understanding).

In addition, each study included several different specifications and therefore for most subjects there were different question paper styles to allow for comparisons and control. For this to be effective, the particular units of the examination selected had to be assessing broadly the same range of knowledge and skills. This gave each study greater focus, but means that the findings cannot be generalised in terms of assessment in the subject.

Each study also offered a good deal of flexibility in the way it developed. For example, at the initial meeting, reviewers sometimes identified specific materials that

would improve the validity of their judgements. These ranged from support materials to increasing the number of series studied.

Phase two – centre visits

In addition to the general limitations of this phase of the work described above, a number of practical difficulties meant that the number and range of centres where teachers and pupils were interviewed were less than expected. Six centre visits were planned for each subject involved in this phase of the work, but for three subjects the number of centres was between two and four.

Both teachers and groups of up to six pupils were interviewed. In the majority of cases interviews took place within centres where pupils had remained in education; ie for the GCSE and AS level units, pupils interviewed were those who had continued their studies, not necessarily in the subject of the predictability study. This meant that it was not possible to include 11–16 schools in the GCSE studies. This rationale allowed interviews to take place with pupils who had been prepared for the summer 2007 examination in the relevant unit, and relatively shortly after they had taken the examination. For A2 units, this approach was not possible. Here, the pupil interviews took place at university departments rather than within a centre. This meant that interviewees were from a number of different centres as well as those who had taken different specifications. This proved to be useful in that the experiences of pupils from different centres studying different specifications could be discussed and compared.

Phase two – desk-based research

This was the area where there can be least confidence in the work. This is partly because it had not originally been designed into the process but was a response to how the work was developing in particular studies. It was also wholly dependent on the availability of item-level data, which was true for only a small selection of the possible units.³ Most importantly, it depended on the aspects of predictability under consideration being identifiable in the statistical information about specific questions. The hypothesis that the more predictable questions would either prove unexpectedly easy or would produce unusually bunched marks, or both, seems reasonable. But it is untested, and even if true depends on accurate estimation of the intended difficulty of a question and/or the extent to which it should spread out marks. It is also worth pointing out that any abnormal statistical patterns might arise from other effects as well as the predictability of the question.

³ It depends on the relevant papers having been marked on-screen or online. Currently, only Edexcel routinely marks in this way.

In the event, it was not surprising that this proved the least informative part of the project, although it provided useful insight into the use of item-level data for future work.

Phase three – review of candidate work

There are several limitations arising from any review of candidate work. It is always limited in scope and greatly dependent on the scripts used providing the appropriate evidence. It is also often the case that the actual answers that candidates write or aspects of the way the answers have been marked become more interesting than the question being investigated.

For these studies several steps were taken to mitigate these effects. First, at the meeting to review the findings of phase one, there was considerable discussion of the nature of work which would best enable further investigation of any issues identified. Then, as already noted above, further care was taken to ensure that the work available was as representative as possible while meeting the agreed requirements. Even so, the work was from whole centres with reasonably large entries, meaning that work from only a limited number of centres was seen, and none from those entering small numbers of candidates in the subject.

Of course, short of asking large numbers of suitably qualified people (and in most cases, there simply aren't large numbers of such people) to take a great deal of time to look at a very extensive range of work, there is no solution to this. In this study, reviewers were looking at scripts in a very focused way. For example, they may only have been looking at how candidates answered specific questions; or they may have been looking for particular patterns of argument in answers from a particular centre. As a result, they were able to consider large numbers of candidates' answers. More importantly, since they were essentially testing a hypothesis, they only had to look at enough evidence to decide whether or not the hypothesis was proven.

There is one exception to this. In some cases, the suspicion was that the particular structure of the question paper meant that candidates would not be taught significant elements of the subject content. Here, reviewers were essentially trying to prove a negative. While this is impossible, it was possible for reviewers to come to an informed view based on the range of answers and approaches within a centre. It did, however, mean that the answer to the hypothesis had to be not proven rather than a confident yes.

Findings

Study 1: GCSE history

The particular focus of this study was on the source-based questions. There had been concerns in scrutiny work and other reviews that these questions did not actually test the skills of critical interpretation of sources as well as they should. (This view was echoed in a report by the Historical Association in 2002,⁴ which expressed particular concern about formulaic questions.) Phase one of the work therefore considered the papers focusing on sources from three GCSE examinations. As a result of this work, the reviewers concluded that there were definite discernible patterns around the nature of the assessment.

The key difference identified was that one of the examinations called for answers on two areas of content which essentially followed exactly the same structure, while the others tested rather different skills across the two content areas. Part of the underlying cause of the differences lay in the way the examination was structured in terms of candidate choice. Where candidates are given free choice in the paper across all topics, it is likely that all questions will have a common structure, since that is a straightforward means of ensuring that all questions address the relevant assessment objectives consistently. Where candidate choice is constrained in some way, as it was with the other two examinations, it is possible to ensure appropriate coverage of the assessment objectives in a more flexible way.

It was also noted that one of the examinations contained a distinctively unpredictable question. These differences made it possible that the two examinations would reward contextual knowledge rather differently. The report also noted that the mark schemes seemed to strengthen this possibility.

This study therefore warranted further investigation, and it was agreed to proceed with both the examination that used entirely predictable structures for every question and the one that contained the unpredictable question.

For various reasons, it proved difficult to arrange centre visits in this study and these therefore concentrated on the examination identified as most predictable. Even here it was possible to arrange only two visits, which means that they provide very limited evidence of substance. However, both centres confirmed the view that teaching and learning, especially in the later stages of preparation for the examination, would pay close attention to the exact structure of the questions and, at one centre at least,

⁴ Tillbrook, M. (2002) 'Content restricted and maturation retarded? Problems with the post-16 history curriculum', *Teaching History*, no. 109.

supply helpful phraseology to answer specific questions. The visits provided some evidence that the topics to be tested in the examination were also relatively predictable, but this had little impact on teaching and learning.

In the script review, evidence from the candidates' answers to this examination tended to confirm the overall concern, although the effect was not a strong one. In particular, the examination makes relatively little demand for contextual knowledge to extend responses to the sources. Such knowledge is highly rewarded as a result. A particular concern about a predictability issue of this type is that candidate performance is more affected by the specific teaching they have received than is desirable. There was some evidence to confirm this, with some centres clearly having prepared candidates much more successfully than others.

Study 2: GCSE media studies

The concern with the particular GCSE media studies examination that was reviewed was that the questions asked about (entirely predictable) content were highly formulaic and very similar across the different examination series.⁵ What is more, the examination is tiered and there was very little difference found between papers at foundation level and those at higher tiers (essentially a one-word difference encouraging some higher-order skills). It seemed probable that this might lead to very formulaic teaching with the content taught only in terms of enabling candidates to provide schematic answers to the predictable questions.

An interesting feature of phase one was that, with no warning so far as reviewers were able to judge, the awarding body changed the nature of the higher-tier questions and changed specific terminology used in both tiers. To expert eyes, the change was seen as minimal: in most cases the quotation which was now used to introduce the question was little more than a distraction. However, it was clear that this added particular interest to phases two and three, in that reviewers were interested to see the effect of this new feature on centres' attitudes to the paper and the way candidates had responded to the new question format.

The centre visits provided little clear evidence about the issues. Candidates were understandably not worried by the generally predictable nature of the question paper, since it gave them confidence that they would receive the right result. On the whole the candidates involved in the centre visits were high attaining and, as will be discussed below, the change to the higher-tier paper did not seem to affect them significantly. Teachers expressed some concern about the paper, feeling that it stifled creativity.

⁵ The choice of content in the question paper, and thus its limitation, is legitimate because it is explicit in the specification, which is clear about the extent of subject matter which candidates are expected to cover. The question paper merely reflects this.

The script review raised some important issues. In general, the changes had had a differential effect. For the more able candidates (such as those involved in the centre visits) the addition of a quotation had been, if anything, enabling. It is a familiar question type in other subjects, and good candidates could transfer the necessary skills to the new context. Less able candidates entered for higher tier were, however, less able to cope with the new question type.

The change of terminology ('representation' had been used in past papers – this was replaced in both tiers with 'messages and values' in 2007) was more problematic. There is no doubt that the concepts are central to media studies, and the phrase occurs in the specification and the heading of the examination paper. However, it seems to have led to considerable confusion among candidates. This is reflected in the chief examiner's report. What this perhaps suggests is an issue arising in several of the studies: the danger of making small changes in largely familiar or even predictable examinations. In this case, it seems to have led many candidates to feel that they needed to present their material very differently from how they would had the question still referred to 'representations'. It seems quite possible that some candidates would have fared better had they ignored the question altogether and answered with what they had prepared about representations, and that cannot be effective assessment.

Study 3: GCSE modern foreign languages

GCSE modern foreign languages have been a focus of much attention lately. As well as concern that they might be harder for learners to achieve than GCSEs in other subjects, there is a widely held view that some of the problems in the examinations derive from the constraints that surround them, leading to very formulaic questions and answers. It seemed sensible to include the subjects in this set of studies.

The decision was taken to look at two languages, French and German, and the writing and speaking components only. From the outset, it was agreed that it would be unlikely that the review of the speaking tests would be extended to phase three, largely because of major logistical difficulties in obtaining and then evaluating the relevant materials.

For speaking, there was evidence of a high level of predictability in the speaking assessment in terms of the topic coverage in the various role-plays, as there was in the writing papers. There was variation in the extent of the predictability across the awarding bodies, with the narrative role-play from one seen as very rigid and unchanging.

One of the problems seen in the writing papers was the effect of the grade description for C. The need for candidates to show use of past, present and future tenses, together with expression of opinions with justification, led to very formulaic and sometimes highly artificial questions. This was notably underlined by the fact that

where there was choice in the writing tasks, the formats were actually very similar. For one awarding body, this was exacerbated by mark schemes which tended to reward safe answers over those which tried a more original approach.

These findings meant that this study was carried on into both phases two and three, focusing on the examination where the issues had seemed most pronounced. The centre visits confirmed a number of the views formed in phase one (and some of the wider concerns about modern foreign language examinations expressed in, for example, the Dearing *Languages Review* report of March 2007).⁶ Both learners and teachers recognised the highly predictable and formulaic nature of the examinations and understood the reasons behind it. Both groups also appreciated these aspects in terms of making preparation for and performance in the examination easier to manage, while finding the process narrow and rather frustrating.

Phase three of the study focused only on the higher-tier writing papers, since it was only these papers which provided the sort of continuous writing which would reward scrutiny. Evidence from this part of the review confirmed that the factors identified at the outset were having the negative effects anticipated. It was particularly striking in the range of vocabulary seen in the answers to the question targeted at grades C and D. This was very limited across all the work seen, and sometimes the same very limited range of descriptive adjectives appeared in the answers from all candidates within a centre. Expressions of opinion were also very formulaic.

There are some significant concerns here, and although they were more severe with one awarding body than the others, it must be stressed that they probably owe more to restrictions imposed by the requirements of the criteria and grade descriptions. The current review of GCSE is looking widely at the examinations to try to address these issues.

Study 4: GCE English literature

Recent work on GCE English literature, especially the standards review in 2006,⁷ had suggested that there were some ways that features of the examinations combined which might have implications in terms of predictability. Accordingly it was decided to include the subject in the project.

Two separate aspects of the assessment had been identified, in both cases when seen in the light of the specific assessments being open book. The first concerned whether, at AS level, essay questions on Shakespeare which were both relatively

⁶ DfES (2007) *Languages Review*, ISBN 978-1-84478-907-8.

⁷ QCA (2007) *Review of Standards in A level English Literature, 2001–2005* (QCA/07/3100).

predictable and limited in scope might lead to poor assessment. The second concerned the way that synoptic skills were assessed in some specifications.

In the AS Shakespeare unit, there were significant variations across the specifications in the extent to which the questions were predictable. This was in terms of both the style of the questions and their content. One examination involved extract-based questions (which was especially doubtful as a question type for an open book examination) and very specific questions on either a single character or a single theme. It was noted that, over time, elimination of passages and major characters would make it easier to predict the questions. This was exacerbated by the explicit opportunity for candidates to limit their answers to two scenes from the play. Interestingly, the paper in summer 2007, although containing some questions that had been anticipated, departed from the pattern of previous papers somewhat, for example offering only one extract-based question.

The A2 synoptic units were notable for the variety of approaches taken. In most cases, the way the examination was structured was judged not to risk predictability. In fact, in some cases, the examinations' unpredictability was judged to make them highly demanding. In one case, however, the assessment was based on pairing plays for study. The reasons for the pairings (which might be thematic or in terms of genre) were not made explicit but were relatively easy to infer, and the plays formed the basis of a substantial element in the examination. It was agreed that the combination of this with the support materials provided by the awarding body, as well as the open book nature of the examination, meant that it might affect what the examination was really testing.

At the end of phase one, it was agreed that the nature of the issues meant that there was little to be gained from centre visits, but that looking at candidates' scripts would help give definition to the issues.

At the script review on the AS unit, reviewers agreed that candidates' responses, in terms of both pattern of choice and the skills shown, did not reflect the concerns. For example, where numbers of candidates from a centre included similar content in response to a particular question, this was not beyond what would be inevitable when answering the question and was countered by the range of arguments presented. Reviewers could see no evidence of either formulaic phrasing of points or structuring of arguments.

In the A2 unit, the outcome of the script review was rather different. The question asked on the overwhelmingly popular option in the paper was multi-layered and complex. However, it started with the key idea which links the texts together (the idea of the tragic hero) and produced many answers which were little more than regurgitations of prepared material. (This was confirmed by many of the comments in the Chief Examiner's report.) Only in answers at the very top of the mark range was

there evidence of candidates showing the full range of expected skills and understanding, and engaging with the question. Reviewers noted that the relatively generous time allocation for the paper exacerbated the problem, with candidates substituting quantity for quality. This is another interesting example of the point that predictability does not necessarily make an examination easier. Here, there was evidence that some candidates had done worse than the general quality of their understanding suggested they should. This was a result of too great a focus on a particular issue in the preparation for an examination, a question that unintentionally (but foreseeably) encouraged that focus, and the availability of time and text to allow candidates to take themselves further and further off target.

Study 5: GCE geography

Geography sets particular problems for assessment. As well as there being a need to assess the familiar range of knowledge, understanding and analytical skills, there are also important geographical skills that candidates must develop. Many of these skills are developed through fieldwork which may be internally assessed, but they are also assessed through external examination. One of the problems that this creates is that the experience on which this assessment is based is necessarily specific to the candidate – or at least the school or college the candidate attends. As a result the questions tend to be very general, which significantly reduces the range of questions that can be asked: and raises the risk of predictability.

In phase one, reviewers were able to identify a range of factors which would increase the level of predictability in a skills-based examination paper and to agree how these factors related to the categories used in the relevant forms. However, this process was not easy, which usefully served to illustrate the extent to which those categories are complex constructs. This suggests the need for refinement of the process were the exercise to be repeated. As a result of the analysis, papers from three awarding bodies were identified as being more predictable than the others. It was also agreed that the nature of the issues meant that while there would be little to be gained from visiting centres, it would be useful to see candidates' work to see how the issues carried through.

The script review looked at three examination papers identified as being high risk in terms of predictability. In each case, the underlying risks identified during phase one and the findings from the review of candidate work differed, so each will be reported separately.

In one case, there were concerns about two separate aspects of the examination paper. It had a narrow content base, restricting the range of coverage possible and making the knowledge-based elements of the paper susceptible to question spotting; and the questions in the skills-based section of the paper were predictable. The scripts seen did not really bear these concerns out. However, this is not the whole story.

Reviewers found that performance on the fieldwork questions was relatively poor given how the candidates had achieved on the remainder of the paper. This suggested that candidates had certainly not been over-prepared for it. Rather, there was little evidence that teachers had done anything to prepare candidates for that section of the paper at all. It was noted that the section carried relatively little weight in the assessment overall, and it seems that it was being neglected. This rather defeats the purpose of the section and raises serious questions about the validity of the assessment overall.

Reviewers also saw little evidence in scripts of over-rehearsed answers to the knowledge-based questions in the summer 2007 examination. However, they did note that these questions had a rather different balance between knowledge and application than in previous series, which probably explained the finding. This was, to an extent, confirmed by the fact that one question part which used wording very similar to previous examinations did produce more formulaic responses. This finding serves usefully to illustrate the complex balance between predictability and unpredictability in an examination.

The second examination raised slightly different issues. First, the unit and its question paper covered only limited content, resulting in a high degree of predictability; this is increased by the design and wording of the questions. The questions covering fieldwork tended to be generic and formulaic in wording. Finally, the range of data-processing skills was very limited, with some appearing in the papers almost every series (and carrying quite high marks for relatively straightforward questions).

For the first of these issues, reviewers did find that candidate work showed evidence of mechanical and formulaic responses. This, coupled with the way that marks were allocated, meant that the questions were not really addressing the assessment objectives as intended. The possible predictability of the fieldwork questions was judged not proven, but the mark allocations meant that they did not contribute sufficiently to the effectiveness of the paper in terms of its discrimination. The main finding, however, was that the way the data-processing skills were assessed did indeed distort the examination, with candidates gaining high reward for an essentially routine (and essentially mathematical rather than geographical) task.

For the third examination reviewed, the fieldwork issue was broadly the same as for the other two. There were also concerns that the questions were generic and, as a result, narrow in range. In this case, however, there was no evidence from the scripts to suggest that these issues had a damaging effect on the assessment.

This study raises some important points. One notable feature is that it is not just predictability that affects the validity of a scheme of assessment. In particular, it is important to ensure that skills are given sufficient weight in any piece of assessment

to ensure that they are properly taught. When assessing skills, it is also important that assessment focuses on their application.

Study 6: GCE law

Scrutiny of a GCE law specification (which also looked in some detail at the other one available) suggested that there were aspects of the examination where predictability may be a problem.

Phase one of the study did suggest that there were indeed aspects of each examination where predictability may be a problem. In both specifications, these issues were of two kinds. First, the structure of some question papers meant that there was no need for candidates to cover all the content suggested in the specification. Second, the pattern of questions across series made it relatively straightforward to predict which topics would be covered in which examination. A particular feature of this on one paper was that in one case a more difficult (and – probably consequentially – unpopular) topic was only ever covered in the January examination, while the summer examination consistently contained the most familiar topics. One effect of this is that for candidates who have not been taught the most demanding content, the two examination sessions are different, with the extent of choice reduced in January. It also has the effect of still further reducing the likelihood of that topic becoming popular. In fact, for this unit, reviewers were able to establish such a consistent pattern of topic coverage that it raised the possibility of candidates covering only a very limited range of the intended content (no more than a third) and still being well equipped to deal with the question paper.

The situation was similar for the other awarding body, where the paper comprised five questions of which candidates had to answer two. There are six topics in the specified content for the paper. Here, a candidate well prepared on three of those topics would have no problems with any combination of topic coverage on the paper. This was exacerbated by the fact that the questions on each topic were very general and, as a result, similar across examination series.

These issues meant that the study did proceed to phases two and three. In both cases, phase two was very limited in scope and does little more than supply evidence that can deepen the findings from the other phases.

For the first awarding body, there was no evidence from talking to centres that teaching had been unduly narrow. The script review was also inconclusive.⁸ In legal terms, the concern can only be described as not proven. However, specific questions did prove overwhelmingly popular with candidates in some centres, which might

⁸ It is important to note that in many ways the exercise was trying to prove a negative (ie that candidates had not covered a sufficient range of content) and was thus impossible.

suggest that they had concentrated on a narrow range of content. This is made more plausible in that the very few candidates who had answered a different question did so badly. However, this cannot be proved and was, in any case, a relatively isolated phenomenon.

In the other examination, evidence from talking to centres again suggested relatively wide content coverage in the teaching and learning for the paper, with narrowing occurring, if at all, only in the last stages of preparation. Even so, there was recognition of the highly predictable nature of the examination and its potential for narrowness of coverage.

As with the first awarding body, the script review was inconclusive. There was strong evidence that some topics were much more popular than others, with some hardly ever selected for response. For some centres, answers were on the same three topics for all candidates, suggesting that narrowing of the taught content had indeed occurred. In fact, for one centre all but one of the candidates answered on the same pair of topics. But this was by no means true for all centres reviewed. Here, too, the issue can only be regarded as not proven.

It remains the case for both examinations, however, that the way the paper is structured and the pattern of questions over series would permit significant narrowing of the taught content. That this does not seem to occur on a large scale reflects well on the professionalism of the teachers, but it is not necessarily a desirable feature of the assessment.

Study 7: GCE psychology

Both a recent standards review⁹ and earlier scrutinies had suggested that there may be issues with predictability in psychology. Phase one looked at two specifications from different awarding bodies and concluded that there were significant aspects of one which were over-predictable. There were several separate features of the examination which taken separately might give cause for concern about predictability but which taken together suggested serious problems. First, the way that one of the A2 papers was constructed meant that candidates only needed to cover half the content that the specification suggested should be taught. Second, the questions were worded using wording directly from the specification, making both the range of possible questions very limited and each question relatively unchallenging. Next, questions were never asked which required candidates to display knowledge from more than a single topic area. Finally, there was a fairly predictable pattern to the rotation of topic areas covered, though it was not an exact rotation. These concerns found eloquent confirmation in the predictions made by the reviewers, with all three recording very high levels of accuracy. (One predicted perfectly a set of questions

⁹ QCA (2007) *Review of Standards in A Level Psychology 1997–2005* (QCA/07/3099).

that would allow candidates to answer the whole paper on what would amount to little more than 20 per cent of the subject content.)

One of the possible effects of the structure of the question paper is that teachers could wholly avoid some aspects of the subject content, and visits to schools and colleges (supplemented by interviews with students now in higher education) confirmed that some of the testing content simply is not taught, and none of the centres visited tried to teach the whole content.

The predictability of the topic coverage and the phrasing of the questions was also confirmed in phase two, with several students commenting that they had answered almost identical questions before they took the examination. All this was supplemented by board-approved books which provide model answers and positively encourage narrow coverage, although several centres are so well versed in the examination that they rely more on highly focused teaching materials that they have produced themselves.

All these issues were further confirmed in phase three, where there was strong evidence of very limited topic coverage. Even where there were answers on more than three subsections of the paper, they always occurred in identical combinations, suggesting that teachers might teach different topics, but that they never taught more than three. Only two of the 15 centres seen appeared to teach four subsections to their students rather than the minimum. Moreover, many answers from some centres were highly similar in structure and content. A striking illustration of this is the extent to which over-coached candidates failed to cope with questions which required some manipulation of the material. Here, such candidates simply repeated the taught material, once again illustrating the point that generally predictable examinations often hinder rather than improve candidates' responses. Reviewers commented that with some centres, answers were so similar it almost suggested malpractice in the examination room, were it not for the fact of how easy it was to produce such consistent answers simply by predicting the questions.

One or two of the centres seen did not appear to have taken such a targeted approach to preparing students for the paper. Reviewers noted that, although it was hard to be sure simply from the answers, candidates at such centres seemed to have been disadvantaged. Certainly, there was evidence from the bulk of centres that relatively weak candidates could be so specifically prepared for the examination that they were able to perform quite well. This, of course, is a major aspect of looking at predictability: while teachers may regret the reductive approach they are able to use, neither they nor their learners are likely to complain given that they get good results.

Study 8: GCE physics

Most of these studies have investigated papers which involve relatively open-ended questions and extended writing. It seemed sensible to include also at least one

subject which characteristically assesses candidates' understanding of content by much more structured questions to see what issues, if any, there are for predictability in such papers. GCE physics was chosen for this purpose.

In phase one, reviewers looked at a range of units from AS to synoptic units at A2. They found that the format of the papers was indeed predictable: that is, papers consisted exclusively, or almost so, of structured questions, each testing a specific area of the subject content. In addition, for some areas of content, the questions in different series were very similar. However, because of the way each paper sampled the subject content, such predictability did not significantly affect the demand of the question paper.

It was noted that the narrower the content in a unit, the greater the risk that predictability would become an issue, simply because most of content was tested in every paper. Even with the least content, however, reviewers did not find any of the papers worryingly predictable.¹⁰ Consequently, it was decided not to proceed to phases two or three, although it was agreed to review item-level data from one unit from summer 2007 to see if performance on any of the questions reflected the extent to which it was predictable.

What this study did usefully reveal is the difference between the format of an examination being consistent and thus familiar to candidates and it becoming so predictable that it no longer really tests what it claims to do.

Study 9: GCE religious studies

Previous scrutiny work on religious studies had suggested that some of the question papers were highly predictable. The study looked at philosophy and ethics papers from several examinations to investigate the question.

Phase one produced a number of interesting findings, although for various reasons it was agreed not to extend the work into phases two or three. The main specification which raised concerns had in fact just been significantly amended for first teaching in September 2007, so that it was clearly inappropriate to investigate the impact of any predictability on the preparation for the actual examination. However, there were sufficient similarities across the two versions of the specifications for the work to be extended into a review of item-level data.

One other feature of the approach to marking in the subject was that, in many cases, markers were instructed to allow credit for material contained in answers to part questions where it was relevant for answers to other parts of the question. This

¹⁰ It is also worth pointing out that the restructuring of A levels for first teaching in September 2008 has meant that there is increased content in the new units.

caused some discussion and it was decided that it was probably a fair approach to the particular format of the examination papers under scrutiny. However, it is not an approach that encourages clear and critical thinking in candidates. It would clearly be preferable to use the structure of the question papers and the way questions are worded to make clear precisely what is expected and then adopt a more rigorous approach to the marking.

The key issue identified in looking at the specification which became the focus of the study was similar to findings in GCE law and GCE psychology: the way the paper was structured was such that it was possible to omit significant amounts of the specified content (especially the more demanding elements) and still be perfectly well prepared for the examination. Reviewers noted that the chief examiner's report suggested that there were great variations in the popularity of questions: while this is not conclusive evidence, it does tend to corroborate this view.

In contrast to this, reviewers agreed that the exact wording of the questions varied in ways that were unhelpful. They agreed that this placed too high a premium on examination technique rather than developed knowledge and understanding. This once again underlines the complexity of predictability: it is not simply a matter of formulaic questions being a problem, nor of avoiding them being the solution.

A further feature of predictability that emerged during this study is how what appear to be different questions actually require virtually identical answers. For example, for a given topic, a question asking candidates to evaluate its advantages and disadvantages, a question suggesting the advantages outweighed the disadvantages, and a question suggesting the opposite view are all in effect identical in terms of the response expected.

Overview

As noted at the beginning, the purpose of the project was to study predictability in examinations. It was not to find fault with the actual examinations used, although the examinations were often chosen because there were suggestions that there might be predictable features in some of them.

One important point that did emerge was how important it is not to use a term like formulaic, which has pejorative overtones. It is important that question papers do not contain major surprises from one series to the next, otherwise far too much of what you will be measuring is how well candidates cope with those surprises rather than knowledge and understanding of the material. In fact, there is a good argument for quite severely limiting the language used in questions, so that it is always clear what is being asked for: there were examples in the examinations seen where the marking schemes suggested there was little distinction between questions using 'describe' and those using 'explain' as their command words. A question does not become easier simply because, as a matter of course, questions requiring an explanation always start with 'Explain why'. This is certainly working to a formula but is likely to lead to more consistent assessment.

Conversely, the increasing tendency to rely on the wording of the subject content is an area where the accusation of being formulaic has some basis. It is an understandable enough approach, neutralising any complaint that a question covers material not in the specification, but it is unhelpful. This is especially the case in A2 units which surely should be set at a level which assesses deeper levels of understanding.

Perhaps the main cause of undue predictability arose from the structure of the papers. There is always a tension between the need to sample content and the need to be fair to candidates whose strengths may vary across the different content areas. The key to the problem is where the structure of the paper allows considerably less coverage of the prescribed content at little or no risk to the candidate. There is no problem if the specification makes it clear that candidates are only expected to cover part of the content: it is then easy for users of the qualification to decide if that is sufficient. It only arises if the implication in the specification is that all the content will be studied when this is not really true.

But the main conclusions to be drawn are two-fold. The first one is that the issue of predictability is much less prominent than many seem to believe. This is in line with

findings from a survey of teachers' attitudes to examinations.¹¹ There are some causes for concern, but they tend to be isolated and specific to a particular paper in a particular specification.

The second main conclusion from the work is that predictability seems to become a significant issue only when several of the possible sources combine. Thus, the psychology examination which caused most concern had an examination paper which permitted significant selection of the intended content, questions which were severely limited in range and a pattern of selection of sub-topic that was too regular. Similarly, the problematic modern language examination addressed a limited range of content and used a particularly formulaic approach to the questions in order to meet the specification. Conversely, the history examination which used relatively formulaic questions was seen to produce better performance from candidates and more reliable marking than one which used a more varied approach.

¹¹ *Survey of teachers' views of GCSE and A level question papers in the summer 2007 examination session*, research study conducted for the Qualifications and Curriculum Authority (QCA 2007).

Appendix A - Specifications and question papers

AQA

(3047) (3042) GCSE History Specification B

Conflict in the Modern World: International and British History (Paper 1)

Governments in Action in the First Half of the Twentieth Century (Paper 2)

(June 2003–6 and 2007)

(3657) (3651) GCSE French Specification A

Higher writing test

Foundation writing test

Higher speaking

Foundation speaking

(June 2003–6 and 2007)

(3667) (3661) GCSE German Specification A

Higher writing test

Foundation writing test

Higher speaking

Foundation speaking

(June 2003–6 and 2007)

(5181) (6181) GCE Psychology Specification A

Social Psychology, Physiological Psychology, Cognitive Psychology, Developmental Psychology and Comparative Psychology (Unit 4: PYA4)

(January and June 2005, 2006 and 2007)

(5741) (6741) GCE English Literature Specification A

Shakespeare (Unit 2: LAW2)

(January and June 2005, 2006 and 2007)

(5061) (6061) GCE Religious Studies

Religion and Human Experience (Unit 1)

(January and June 2003–6 and 2007)

Religion and Ethics (Unit 4)

(January and June 2003–6 and 2007)

(5031) (6031) GCE Geography Specification A

Geographical Skills (Unit 3: GGA3)

(January and June 2005, 2006 and 2007)

(5161) (6161) GCE Law

Law Making (Unit 1: LAW1)

Criminal Law (Offences against the Person) or Contract (Unit 4: LAW4)

(January and June 2005, 2006 and 2007)

(5451) (6451) GCE Physics Specification A

Particles, Radiation and Quantum Phenomena (Unit 1)

(January and June 2003–6 and 2007)

CCEA**GCE English Literature**

The Study of Shakespeare (Module 2: ASL21)

Drama (Module 6: A2L31)

(January and June 2005, 2006 and 2007)

GCE Geography

Techniques in Geography (Module 3: ASG31)

(January and June 2005, 2006 and 2007)

GCE Physics

Forces and Electricity (Module 1: ASY11)

(January and June 2003–6 and 2007)

Edexcel

(3334) (1334) GCSE History (Modern European and World History)

Depth Studies (Paper 2)

(June 2003–6 and 2007)

(3226) (1226) GCSE French

Writing – terminal examination (Paper 4F)

Writing – terminal examination (Paper 4H)

Speaking – terminal examination (Paper 2F)

Speaking – terminal examination (Paper 2H)

(June 2003–6 and 2007)

(1231) GCSE German

Writing – terminal examination (Paper 4F)

Writing – terminal examination (Paper 4H)

Speaking – terminal examination (Paper 2F)

Speaking – terminal examination (Paper 2F)

(June 2003–6 and 2007)

(8180) (9180) GCE English Literature

Criticism and Comparison (Unit 6: 6396)

(January and June 2005, 2006 and 2007)

(8562) (9562) GCE Religious Studies

Philosophy of Religion (Unit 2: 6772)

(June 2003–6, January 2007)

Religious Ethics (Unit 3: 6773/02)

(June 2002–6, January 2007)

An Introduction to Religion and Ethics (Unit 4: 6581)

(June 2007)

(8214) (9214) GCE Geography Specification A

Applied Geographical Skills (Unit 3b: 6463)

(June 2003–6 and 2007)

(8540) (9540) GCE Physics Specification A

Mechanics and Radioactivity (Unit 1: 6731)

Synthesis (Unit 6: 6736)

(January and June 2003–6 and 2007)

OCR

(1937) GCSE History Specification B – Modern World

British Depth Study (Paper 2)

(June 2003–6 and 2007)

(1925) GCSE French

Writing/Foundation (Unit 4: 2354F)

Writing/Higher (Unit 4: 2354H)

Speaking – externally assessed/Foundation (Unit 2: 2352F)

Speaking – externally assessed/Higher (Unit 2: 2352H)

(June 2003–6 and 2007)

(1926) GCSE German

Writing/Foundation (Unit 4: 2364F)

Writing/Higher (Unit 4: 2364H)

Speaking – externally assessed/Foundation (Unit 2: 2362F)

Speaking – externally assessed/Higher (Unit 2: 2362H)

(June 2003–6 and 2007)

(1918) GCSE Media Studies

Cross-media topics – Foundation (Unit 5: 1918/5)

Cross-media topics – Higher (Unit 6: 1918/6)

(June 2003–6 and 2007)

(3876) (7876) GCE Psychology

Psychological Investigations (Unit 2542)

(January and June 2005, 2006 and 2007)

(3828) (7828) GCE English Literature

Drama: Shakespeare (Unit 2707)

Comparative and Contextual Study (Unit 2713)

(January and June 2005, 2006 and 2007)

(3877) (7877) GCE Religious Studies

Foundation for the Study of Religion (Unit 2760)

Philosophy of Religion 1 (Unit 2761)

Religious Ethics (Unit 2762)

(June 2003, January and June 2004–6 and 2007)

(3832) (7832) GCE Geography Specification A

Geographical Investigation (Unit 2682)

(January and June 2005, 2006 and 2007)

(3839) (7839) GCE Law

Criminal Law 2 (Unit 2572)

Law of Contract 2 (Unit 2575)

(January and June 2005, 2006 and 2007)

WJEC

GCE English Literature

Shakespeare (Unit 1: ELit1)

Drama – Pre-1770 and linked material (Unit 6: ELit6)

(January and June 2005, 2006 and 2007)

GCE Geography

Investigative Geography (Unit 3: GG3a)

(January and June 2005, 2006 and 2007)

GCE Physics

Waves, Light and Basics (Unit 1: PH1)

(January and June 2005, 2006 and 2007)

Appendix B – Contributors

Awarding bodies:

AQA

CCEA

Edexcel

OCR

WJEC

Consultants:

English Literature

Russell Carey

Mick Connell

Alison Woolard

Geography

Glennis Copnall

Sue Driver

John Vernon

History

Ruth Tudor

John Warren

Law

Peter Darwent

Mark Wilson

MFL

Mary Culpan

Bridget Smith

Peter Willig

Physics

Dave Kelly

John Skevington

Nick Cox

Psychology

Diana Dwyer

Mike Kilbride

Andrew Windass

Religious Studies

Libby Ahluwalia

John Rudge

Anthony Glachan

John Summerwill

Media Studies

Richard Hoyes

David Lewis

Centres:

Arnold Hill School & Technology College, Nottinghamshire

Finham Park Comprehensive School, Coventry

Heathside School, Weybridge

Hitching Boys' School, Hertfordshire

Ilkley Grammar School, West Yorkshire

Little Heath School, Reading

Ludlow Church of England School, Shropshire

Newcastle College, Newcastle

Newham Sixth Form College, London

Portland School, Worksop

Royal Grammar School, Newcastle

St Crispin's School, Wokingham

Stockton Sixth Form College, Stockton-on-Tees

Sydenham School, London

Weald of Kent Grammar School for Girls, Kent

West Nottinghamshire College, Mansfield

Universities:

Durham University

Nottingham Trent University

University of Nottingham

Archived Content

Appendix C – Consultants’ templates

Phase one – materials review

Name:

Qualification and subject: GCE/GCSE [Subject]

Series covered: January/June [XXXX], January/June [XXXX], January/June [XXXX], January/June [XXXX]

Awarding body and unit: AB/Level/Subject/Unit

KEY:

Very unfamiliar				Appropriate					Very predictable	
1	2	3	4	5	6	7	8	9	10	11

For each of the following issues, please give a rating using the above key, and comment regarding this.

Issues:

Relationship between specification content and question paper

Level of definition of question paper in specification

Nature/format of questions within a series

Nature/format of questions across series

Availability and nature of other sources of guidance for preparing for examination

At the end of the form, on the basis of your judgements and comments please state your prediction for the summer 2007 question paper.

1. Relationship between specification content and question paper

Rating (between 1 and 11)	
------------------------------	--

Comment:

2. Level of definition of question paper in specification

Rating (between 1 and 11)	
------------------------------	--

Comment:

3. Nature/format of questions within a series

Rating (between 1 and 11)	
------------------------------	--

Comment:

4. Nature/format of questions across series

Rating (between 1 and 11)	
------------------------------	--

Comment:

5. Availability and nature of other sources of guidance for preparing for examination

Rating (between 1 and 11)	
------------------------------	--

Comment:

On the basis of the above, what is your prediction for the summer 2007 question paper?

Phase three – review of candidate work

Name: Consultant

Candidate Scripts

Responses Sheet

Specification: GCSE/GCE [Subject].....

Awarding body:

Issue X

.....

.....

.....

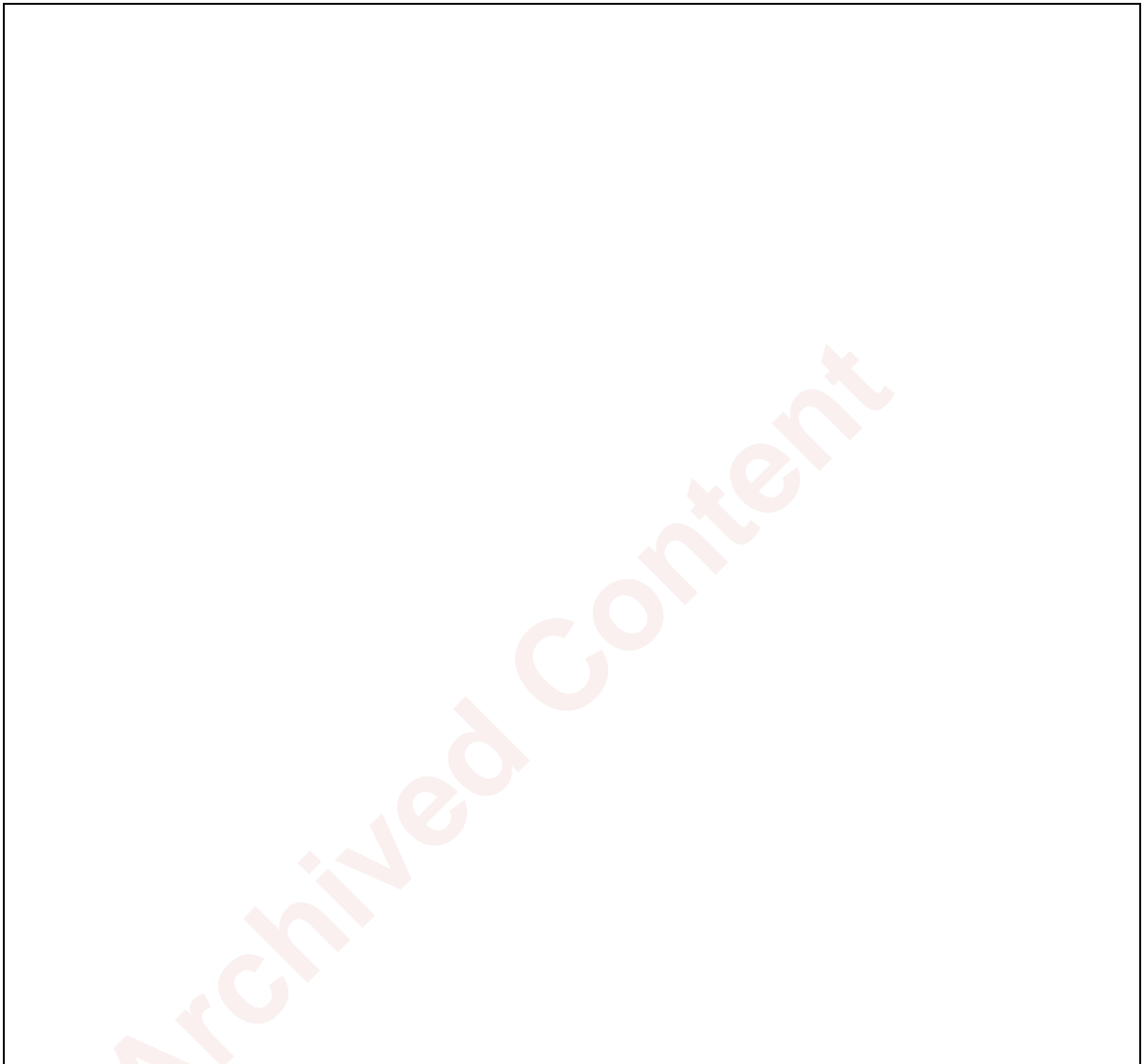
Issue X

.....

.....

.....

A: Summary: overall findings



B: Any other comments overall



C: Summary: Issue X

[Issue X]

.....
.....
.....

You will need to outline all the evidence used to:

- either support or dismiss the issue
- state how easy it was to come to that view
- give specific examples of the particular features raised.

Archived Content

D: Candidate work

Work was considered from the following centres:

Seen (✓)	Centre name(s)/number(s)	Total number of entries	Range of marks seen eg 23, 3x26, 28, 2x35

Any other material used as evidence: e.g. item-level data, examiner's report

E: Notes

As you work through the candidate scripts centre by centre first use the space below to record notes about **your judgements** and **evidence for those judgements** for the predictability issues.

It may not be necessary to look at every script within the centre for you to come to a judgement. Therefore it is important for you to note how easy/difficult it was to come to that judgement and to note the evidence you have seen, i.e. the range of scripts that you looked at, issue was only evident within certain centres/certain mark ranges etc. Whenever possible, cite specific examples (e.g. Centre X, Candidate y, q, z, followed by 'quotation').

After looking at all the centres necessary for this issue you will need to summarise your notes for each issue at the end of this form.

Appendix D – Common schedule of questions

Phase two – centre visits

Suggested structure for questions:

After the pupil has looked at the question paper. The words in italics under each sub-question are prompts to be used if the pupil does not give a good answer.

Questions to pupil

Area 1:

Whilst looking at the paper

1. Did you feel that you were well prepared for this exam?

What was good? What was bad? E.g. question type, structure of paper, topics that came up.

Prompt questions if they are not sure what to say, for example:

- *What went well?*
- *Which questions did you do on the paper if there is a choice?*
- *Why?*
- *Were there any questions that were harder than expected?*
- *Why?*
- *Was/was not covered in the course?*
- *Did/did not understand what the question was asking?*
- *Ran out of time?*
- *Were there any questions you felt really confident about answering?*
- *Why?*
- *It was covered well in the course?*

- *The topics/areas were expected to appear in the paper?*
- *It was a question that they expected to see/similar to something that they have done before as part of their course?*
- *Are there questions that you thought might be on this paper that did not appear?*
- *What did appear that you expected?*

Area 2:

2. How much did revision techniques or guidance from your teacher help you prepare for this exam?
3. How? Why?

Prompt questions, for example:

- *What did you use to help you prepare for this paper? What would you say was most useful?*
- *Past papers?*
- *Notes from class materials?*
- *Internet sources/revision websites?*
- *AB materials/AB training day?*
- *Parental help?*
- *Teacher directed topics expected to appear in the paper?*
- *Teacher guidance on command word responses?*
- *Revision based homework?*
- *Practice questions?*
- *Intranet materials?*

- *How did you know how to structure your answers?*
- *Past papers?*
- *Question types?*
- *Model answers?*
- *Did you use past papers? Were they useful? How?*
- *Were you told how they might help you in this year's exam?*
- *Did they help you to choose topic areas to revise?*
- *How well do you think your teacher prepared you for your exam?*

Area 3:

4. Do you think that you were as prepared as you could be? Was the paper fair?

Prompt questions, for example:

- *Do you think that your course was taught well? Did it cover all the things that you needed to take this exam?*
- *Why?*
- *Did you put in more revision time on a particular topic(s)?*
- *Which ones and why?*
- *Did your teacher suggest topics to revise?*
- *Do you think the examination was a fair test?*
- *Why?*
- *Past papers were very different?*
- *Questions on the paper covered areas not taught?*
- *Topics on the paper were the only topics taught?*

- *Do you think you or your school could have done more to prepare you for this exam?*

Briefing for: Questions for teachers on centre visits

Area 1:

1. How did you go about preparing your students for this exam this year? Was it any different from previous years?

Prompt questions, for example:

- *Do you cover the whole syllabus in your course?*
- *If not, why? How do you choose what to cover?*
- *When do you start preparing your students for this exam? How do you go about it?*
- *What type of revision do your students undertake for this exam? How is this structured in the overall course?*
- *How much do you refer to past examination papers? Do you incorporate this into class time? How?*
- *Do you also create your own test material in preparation for the examination? If so, please give details.*
- *Do you concentrate your teaching in certain areas of the syllabus in the run up to the examination? Why do you think that this focus is necessary?*

Areas which may also be covered:

- Setting revision homework?
- Practice papers?
- Practice questions?
- Revision classes/extra lessons?
- School/department intranet materials?
- Devising revision timetables?

- Providing model answers?
- Helping/recommending to choose areas to revise?
- Exploring command words (such as Explain/Analyse etc.) and how to deal with different types of question?

Area 2:

2. Do you think that your students were as prepared as they could be based on the materials available to you?

Prompt questions, for example:

- *How do you think candidates found the paper? Did your students report anything unexpected?*
- *What?*
- *Do you think the results of the examination reflected the ability of the students in your entry?*
- *Do you think the syllabus is suitably varied to provide wide learning about the subject for your students?*
- *Do you think that the examination papers cover enough of the syllabus over the years? Is it a balanced coverage? Do you have any comments on the syllabus coverage by the examinations?*

Areas which may also be covered:

- From their course are students likely to be able to attempt/answer questions on the paper?
- Were there any completely unexpected topics on the paper?
- Was the structure of the paper similar/same as past papers?
- Were there any questions in particular that the pupils had real difficulty with? And why?

Area 3:

3. How effective were the syllabus and AB support materials in helping you teach the course and prepare your students for their examinations?

Prompt questions, for example:

- *Are there any suggestions that you may have to improve the awarding body guidance for the subject?*
- *To what degree were you able to help your students decide which topics/areas that they should consider revising for their examinations based on the AB guidance and materials available?*
- *With comparison to the past 2/3 years, how easy was it for you to prepare students for this year's papers? Please explain your reasons for this answer.*

Areas which may also be covered:

- Did their students use any AB support materials?
- AB teacher guides/unit guides?
- Information on websites?
- Past papers?
- Outlining what is required of command words (describe, evaluate etc.)?

Further subject-specific questions

MFL questions

Teachers

- Are there any significant issues between the tiered papers in terms of the style of the questions/the way that papers are structured? Are candidates able to prepare for this?
- What do you think about the choices of questions that the students are presented with within the exam?
- Are students able to illustrate their French/German vocabulary on the question papers? Do the papers require them to use more vocabulary or less than indicated in the syllabus?
- What do you think about the awarding body guidance materials?

History questions

Teachers

- Are you happy with the range of skills your students are asked to demonstrate in this examination?

Media studies questions

Teachers

- Were all your students able to access questions in the papers whether they took the foundation or higher tier?
- Did any of your students report any difference in this year's paper compared to previous years? Did your students experience any particular problems?
- How do you help your students to understand what is required of them from different types of questions?
- Do you think that the examination is a good test of the knowledge and skills you would like to see developed in your subject?

Psychology questions

Teachers

- What AB recommended textbooks, Inset, resources did you use/attend to help you teach the course? How valuable were they?
- Did you use the AB Inset and/or AB recommended textbooks? How helpful did you find them?
- Do you know what questions your students attempted on the paper?
- When are students assessed for this unit? January/June/both?

Students

- What AB recommended textbooks, course, resources did you use/attend to help you on your course? How valuable were they?
- How many 'sections' of the paper/questions are you prepared for?
- Was there a difference in the number of topics you were taught about and the number you revised? Why?

Appendix E – Subject issues

GCSE French and German

The vocabulary and structures required are very restricted and repetitive.

- The higher-tier narrative is restricting and allows for only very limited creative language. At higher-tier level more opportunities for creativity should occur.
- The unchanging format of the examination is restrictive per se, even at higher-tier level, compared with the other awarding body specifications.

The situations/contexts of tasks are highly predictable and show very little variation over the years.

- The contexts are inauthentic in a lot of cases.
- The higher-tier tasks are too predictable compared to the other awarding bodies.

The approach of the specification encourages candidates to regurgitate language rather than deploy language creativity compared with the other two awarding bodies.

- There is a very limited opportunity for some use of creativity in higher-tier writing in both French and German.

GCSE history

Impact of mark schemes on reward given to content?

- Do scripts show evidence that contextual content is under-rewarded? NB. This may be dependent at different points of the mark range.

Do formulaic questions lead to candidates answering in very formulaic ways?

- Answers within a centre over-consistent in structure?
- Questions answered across a centre show evidence of working to a formula (same structures across all sections of the paper centre by centre)?

The examination is less predictable in terms of structure, but it may be possible to predict content.

- There may be centre effects in terms of the content used.
- Candidates may struggle to interpret questions and use inappropriate content or in response to wrong question point.

- Candidates may be rewarded for style and fluency rather than content.

GCSE media studies

The foundation paper in terms of the structure and the question wording allows centres to specifically prepare candidates for this paper. Questions are repeated almost word for word. In June 2007 the higher-tier paper changed. Has the 'unpredicted' use of a quote in Questions 3 and 4 of the summer 2007 higher-tier paper appeared to have caught out candidates unfairly?

- What are the differentiating characteristics of candidates at different grades/tiers? Is there noticeable differentiation between the higher- and foundation-tier papers (the news vs advertising questions)?
- Do candidates in the higher-tier paper struggle to interpret the questions and use inappropriate content or respond inappropriately to the wrong question point?
 - Do foundation-tier candidates struggle with the lack of scaffolding?
 - Do foundation-tier candidates show evidence of carefully prepared answers?
 - Do higher-tier candidates struggle to make use of the quotation, or do markers/candidates show any signs of having expectations arising from it?

GCE psychology

Centres exploit the structure and predictability of the paper to prepare candidates to answer perhaps three or four questions, using past papers as a main resource.

- Evidence of strong similarities in the responses of candidates from the same centre to the same question:
 - choice of the same questions
 - the same structure and approach
 - the use of the same examples
 - references to the same experiments
 - the use of common phrases and expressions
 - widespread use of the model answers in the set text.

GCE English literature

The issues addressed in questions – both general essay and extract-based – tend to be fairly obvious ones (eg central characters and major scenes): the range of potential issues will decrease over time as these are used.

- Does it appear that the specific questions posed have been anticipated unduly?
- Is there anything to note from the popularity of different questions?

This is an open text examination (but not 'plain text'), so candidates' notes may address issues in questions very specifically (eg the development of character, or the wider significance to the text of particular scenes).

- Do candidates from the same centre respond overall in a similar fashion and/or refer to the same aspects of the text?
- Do candidates from the same centre use the same irrelevant material? (This may suggest that they have relied on teachers' preparation irrespective of the specific question, particularly if this is evident across the whole ability range.)

The general essay questions tend to give candidates the option of focusing on two extracts or ranging more widely [and prior to this series, candidates could opt for a given text to produce a general essay or to respond regarding an extract].

- Are candidates' responses limited in scope or do they address the text as a whole?

The specific pairings of the six texts may suggest likely issues for questions, and the issues addressed in questions tend to be fairly obvious ones (eg major themes such as form and character), owing to the need to be relevant to both texts: the range of potential issues is relatively limited, and will decrease over time as these are used.

- Does it appear that the specific questions posed have been anticipated unduly?
- Do candidates' responses address fully the content of questions or are they less directly relevant to this and more generic?

The question stems are extremely detailed and have become increasingly formulaic, and the question papers also provide extensive guidance regarding the requirements of the different assessment objectives.

- Do candidates' responses address fully the requirements of questions or are they less focused on these and more formulaic?

GCE geography

The limited number of specification topics, together with the pre-release of topic titles, creates a narrow knowledge base for assessment and increases predictability. There are two optional themes, one from the AS Physical module and one from the AS Human module. Candidates select to answer a question in the context of either physical or human.

- Is there evidence that a narrow range of context increases predictability and influences performance?
 - Are they all choosing the same topics?
 - Do candidates of different abilities appear to know these topics equally well?

Generic questions in the fieldwork section of the assessment are highly predictable.

- Are candidates producing formulaic answers to formulaic fieldwork questions?
 - Is there sufficient convincing evidence of reference to real fieldwork and personal experience?
 - To what extent do answers respond to the precise wording of questions?

Inclusion of knowledge-based questions which are little different from a straightforward Physical or Human paper increase the potential for question spotting.

- Is there evidence that candidates have received highly targeted preparation?
 - Obviously practised question responses?
 - May not always fully fit requirements of the question?

Assessment is predictable by topic and question wording. Some questions repeated almost word for word.

- What are the differentiating characteristics of candidates at different grades answering familiar/generic fieldwork questions?
- What evidence is there that candidates practise to perfection?
 - Are the majority of candidates performing exceptionally well?
 - What is the range of marks for particular questions? Does it reflect the range of final outcome? Is the expected mark range narrowed?

Wording of questions on fieldwork tends to be formulaic, open ended and generic.

- How much do candidates responses vary?
- How are they being marked on this?
- Is the expected mark range narrowed?

There are only 6 groups of data-processing skills and techniques in the specification – these appear in the assessments with great regularity.

- Does the candidate work allow for a judgement to be made on whether questions on data-processing skills and techniques are predictable?
- Do candidate responses to the question on Spearman rank and sampling differentiate?
- How do data-processing questions differentiate?
 - For Spearman rank – appears in every paper. Are candidates performing at the top end of the total marks?

Is the perceived predictability in the fieldwork question having an impact on the overall effectiveness of the assessment?

- How does candidate performance differ between Q1 and Q2 in Knowledge, Understanding and Skills?
- How do marks differ? What is typical performance/difference at the different grades/grade boundaries, ie is the difference more or less marked for candidates of different abilities?
- Is there sufficient convincing evidence of reference to real fieldwork and personal experience?
- To what extent do answers respond to the precise wording of questions?

The examination uses a narrow range of open and generic questions.

- Is there any evidence that the narrow range of context for Question 1 has increased predictability?
 - Is there a good spread of knowledge and understanding or are candidates generally performing the same? What separates them?
- Is the full mark range being used for very familiar/predictable generic fieldwork questions?

- For the questions which you correctly predicted, what evidence is there of a difference in candidate performance (ie compared with questions you considered not predictable)?

GCE law

The way in which the topic options appear on the paper year on year has meant that centres are not preparing their candidates for the full syllabus. Candidates across centres can prepare for half the number of topics and be able to complete the paper (6 topics match 5 questions).

- There are strong similarities in the responses of candidates from the same centre to the same questions, across all centres.
- Candidate responses are based on practised prepared answers, within centres, across teaching groups/across centres. (Statutory interpretation questions vary little year on year.)
- Responses may closely follow model textbook/unit guides.
- Candidates pull out source material for answers rather than applying their K & U (eg legal definitions).

Impact of mark schemes on predictability?

- Emphasis of AO weightings are not the same in practice (ie knowledge and skills in candidate responses do not match the AO weightings).

The limited number of topic options that have appeared on the paper year on year has meant that centres are not preparing their candidates for the full syllabus. Candidates across centres are being prepared for only 2 or 3 topics rather than the full 6 topics.

- There are strong similarities in the responses of candidates from the same centre to the same questions, across all centres.
- Candidate responses are based on practised prepared answers, within centres, across teaching groups/across centres (precedent and statutory interpretation especially).
- Knowledge is not always suitable for specific context.

Appendix F – Resources and materials for each phase

Phase one – materials review

- GCE and GCSE specifications
- Copy of four series of materials – question paper(s), mark scheme(s) and examiners' report(s)
- Inset/revision guides

Phase two – centre visits and desk-based research (where applicable)

- Common schedule of questions leading to:
 - face-to-face interviews with teachers and students (sixth formers and undergraduates)
 - telephone interviews with teachers
- Copy of summer 2007 materials – question paper(s), mark scheme(s) and examiners' report(s)
- Item-level data for question papers

Phase three – review of candidate work (where applicable)

- GCE and GCSE specifications
- Copy of June 2007 question paper(s), mark scheme(s) and examiners' report(s)
- Inset/revision guides
- Reports from phases one and two
- Item-level data for question papers

Ofqual wishes to make its publications widely accessible. Please contact us if you have any specific accessibility requirements.

First published by The Office of the Qualifications and Examinations Regulator in 2008.

© Qualifications and Curriculum Authority 2008

Ofqual is part of the Qualifications and Curriculum Authority (QCA). QCA is an exempt charity under Schedule 2 of the Charities Act 1993.

Reproduction, storage or translation, in any form or by any means, of this publication is prohibited without prior written permission of the publisher, unless within the terms of the Copyright Licensing Agency. Excerpts may be reproduced for the purpose of research, private study, criticism or review, or by educational institutions solely for education purposes, without permission, provided full acknowledgement is given.

Office of the Qualifications and Examinations Regulator
Spring Place
Coventry Business Park
Herald Avenue
Coventry
CV5 6UB

Telephone: 0300 303 3344

Textphone: 0300 303 3345

Helpline: 0300 303 3346

www.ofqual.gov.uk