IS AN EXAM PAPER GREATER THAN THE SUM OF ITS PARTS?
A Literature Review of Question Paper Structure and Presentation

Victoria Spalding

1. SUMMARY

Literature relating to question paper structure and presentation was reviewed in order to gain insight into good practice for question paper writing. This issue is particularly relevant at present as AQA have announced plans to revolutionise their exam construction process; moving towards using question banking. There is very little literature available which directly addresses question paper structure and presentation. Nonetheless, some general guidance has been produced from reviewing the existing literature, though it should be noted that hard and fast rules should not be applied to question paper writing. AQA's current practices were discussed in light of the general guide generated from the research, revealing potential areas which AQA may choose to address. The table below summarises the key issues raised by the literature review.

Research suggests	AQA practice	AQA action
Combined answer booklets are preferable to separate answer booklets, although in some cases (for example long answer questions), separate booklets are preferable.	Currently, AQA advises the use of combined booklets, particularly in light of the move to CMI+. Some difficulties with using combined booklets have arisen. Solutions for providing extra answer space are being considered.	AQA is continually reviewing and enhancing CMI+ marking.
Layout of information needs to be considered: placing key information at the top of the page; ensuring the instructions for the task are the last thing candidates read before they respond; taking into account the effect of preceding questions.	The layout of information and the use of stimuli differ across AQA papers. Most AQA Principal Examiners have naturally adopted these rules for writing papers and, as a result, most AQA papers are currently written with consideration to information layout and appropriate use of stimuli.	More guidance to be provided for question paper writers regarding layout of information.
Optional questions are attractive to teachers and students but comparability is nigh impossible to ensure.	Optional questions are frequently used in AQA's papers.	It may be beneficial for AQA to investigate new practices, for example post-hoc scaling methods to improve comparability.

Research suggests	AQA practice	AQA action
The types of question included in the paper should assess the required skills and knowledge, and vary in style/format.	Currently, AQA does not provide question paper writers with procedural guidance for structuring papers. However, unwritten rules of good practice have evolved over the years and most of AQA's papers do naturally follow the guidelines outlined by the research literature. AQA plan on altering their procedures for question paper writing, introducing a question banking process in the future.	It would perhaps be worthwhile for AQA to make these unwritten rules explicit to ensure a more standardised practice across subject departments, as well as assisting inexperienced paper writers. Equally, AQA need to have a good understanding of how their papers will be affected by the new process of question paper writing.
Questions papers should have an incline of difficulty.		
Questions which relate to pre-prepared work should be positioned near the beginning of the paper.		
Exams should be structured in a familiar style to candidates whilst not being overly predictable.		

Due to the lack of research available on question paper writing, the conclusions which can be drawn from this review are limited. More research is required to directly address question paper structure and content specific effects. It is unlikely that a 'one-size fits all' philosophy can be applied to question paper writing. Current processes for test construction need to be made explicit, and a guide to good practice established, in order to ensure that the new process of exam construction using question banks will not adversely affect question paper quality.

2. INTRODUCTION

There has been a surge of interest in question writing in recent years, with a wealth of research produced on how to write individual questions (Case & Swanson, 2001; Chamberlain, 2009; Crisp & Sweiry, 2006; Fisher-Hoch, Hughes & Bramley, 1997; Green, 2003; Haladyna, Downing & Rodriguez, 2002; Hughes, Pollitt & Ahmed, 1998; Johnson, 2004; O' Donovan, 2005; Pollitt & Ahmed, 1999; Pollitt & Ahmed, 2001). However, the matter of how questions are put together to form a paper has not received the same attention. Whilst question writing is a fundamental skill for assessment, advances in technology raise new questions for question paper writing. The dawn of online assessment and the advancements in question banking technology have made on-demand examinations a real possibility (Wheadon, Whitehouse, Spalding, Tremain & Charman, 2009). On-demand could revolutionise high stakes examinations: potentially papers could be put together on-demand, from a question bank, using a matrix of question difficulties and assessment objective information. Papers would no longer need to be written by one examiner; the questions could be written by a team of experts and put together to form a paper by a separate team of professionals in a semi-automated process (Whitehouse, He & Wheadon, 2008). As technology develops, it may even be possible to construct tests in a fully automated process, which would have time and cost benefits. AQA have embraced question banking and plan on changing their processes; in the future AQA's papers will be constructed by a senior examiner using a question bank. Before we change the processes by which our papers are produced, it is worthwhile asking some fundamental questions: does the presentation of the paper make a difference to candidates? Can the structuring of the paper affect performance? In short, is a paper only as good as the questions within it or is it greater than the sum of its parts? This review of the literature on question paper construction will examine the presentational and structural considerations for paper writing. Whilst question writing issues will naturally arise, the

focus of this review is upon format, presentation and structure of papers as a whole (for a discussion of question writing issues see Chamberlain, 2009). This review is chiefly concerned with traditional printed exam papers; however some on-screen research will be discussed.

3. PRESENTATION

This review will first address the question of whether paper presentation affects candidates' performance. A candidate's score on any given paper is the result of interactions between the questions set, the ability of the candidate, and a variety of other factors which may affect a candidate's performance. The function of a question paper is to assess a candidate's ability in a given topic. However, if the presentation of the paper misleads candidates, causing them to respond inappropriately, then can we be sure that the paper is accurately assessing candidates in terms of ability in the given topic?

3.1. Presentation and the effect on readability

In high stakes examinations, candidates are under pressure to perform well and this heightens their test anxiety. Anxiety can lead to candidates becoming easily confused and increases the likelihood of making mistakes in the exam; for example, candidates commonly misread questions in exam situations (Mulkey & O'Neil, 1999). One of the simplest approaches to preventing this is to present the paper in the clearest possible format.

Text size and font have been shown to affect the readability of printed articles. Arial and Times New Roman are the most commonly used fonts in printed material and are therefore most familiar to candidates and the obvious choice for typesetters (Bernard, Chaparro, Mills & Halcomb, 2003). There has been debate in the past about the use of serif and sans-serif typefaces in print-based materials. Serif typefaces, such as Times New Roman, have cross-strokes that project from the main stroke of a letter and have been found by some authors to be more readable by adults than sans-serif typefaces, such as Arial (Sanocki, 1991; Lanbuz, 1988 cited in Bernard *et al*., 2003). However, other studies have not found significant differences between them in terms of readability (Paterson & Tinker 1932; Poulton, 1965). Moreover, it has been argued that Arial and Comic (sans-serif) fonts are more accessible to children as they are similar to their handwriting and less formal than Times New Roman (Bernard, Chaparro, Mills, & Halcomb, 2002). On-screen research provides support for this argument; Bernard *et al*. (2003) found that Arial and Comic typefaces were preferred to Times New Roman for on-screen reading (Bernard *et al*., 2003; Bernard *et al*., 2002). As a rule, larger text sizes are more readable than smaller sizes; however, the size differences between texts often have to be quite large before any significant difference in readability occurs. Size 12pt is considered the optimum size for reading (Bernard *et al*., 2003). By June 2010, all of AQA's papers will be in Arial front size 11pt; previously AQA's papers were printed in Times New Roman. The move to Arial has occurred after the Disability Discrimination Act advised that Arial was more accessible to visually impaired candidates.

The page layout can also affect the readability of the paper. The use of white space is thought to aid comprehension and display the structure of the paper, as well as improve the aesthetics of the paper (Crisp, 2008). However, the amount of white space used needs consideration; Knupfer and Mclsaac (1992) observed a significant drop in on-screen candidates' comprehension when more than half an inch of white space surrounded the image. The authors speculate that this is due to the cognitive effect of 'proximity', where it is more difficult to follow the test and perceive meaning with large gaps that the eye must jump across to continue

reading. It would appear that too little space makes the paper hard to read whilst too much reduces comprehension. Reading papers on-screen is harder than reading from a printed document (Bernard *et al.*, 2003); the amount of white space is likely to have less of an impact upon printed papers. There is currently no upper limit for the amount of white space which surrounds an image in the AQA style guide. The space surrounding the image is based upon practicalities. The image must be on the same page as the question, or on the adjacent page, so that they can be viewed simultaneously whilst being as conservative with the amount of space as possible. Enforcing an upper limit for the amount of space which surrounds an image is impractical for printed papers.

The format in which the information is presented also impacts on candidates. Johnson (2004) interviewed eleven year olds to discover their preferences for an English test. The participants liked bullet points and images which gave information, though this did depend upon the task. The participants in Johnson's study were younger than the candidates taking high stakes exams and we would expect that they need more support than GCSE or A-level candidates. Equally, the format of the response space can affect performance on a question. Research involving different versions of the Alpert-Haber Achievement Anxiety Test demonstrated that varying the response format of the questionnaire yielded different responses from participants. The Alpert-Haber Achievement Anxiety Test is a self-report questionnaire which, by nature, is subjective and may be more susceptible to response format effects (Huck & Jacko, 1974). However, research into 'fill in the blank' style exam questions has also found response format effects. Candidates performed better on 'fill in the blank' questions when dashed lines within the text were used as opposed to when straight lines or boxes outside the text were used to indicate the blank (Hartley & Trueman, 1986).

3.2 Presentation as an indication of the appropriate response

The presentational considerations for paper writers extend beyond readability. Presentation of the paper can affect how candidates interpret the requirements of the questions. The performance of candidates will, to some extent, be mitigated by how well they understand what the questions require of them. If candidates regularly misinterpret the level of detail with which a question intends them to respond then we cannot be sure that the exam is accurately assessing the candidates' knowledge in a given subject.

Exam papers can be presented as either a booklet, where the questions and spaces for the candidate's responses are combined, or as a separate question paper and answer booklet. Crisp's (2008) research showed that, in general, candidates achieved better marks, wrote longer and fuller answers, and were more precise when using combined booklets compared to separate booklets. In interviews, the participants reported that the response space in the combined booklet encouraged them to answer questions and provided guidance regarding the length and depth of response required. Whilst Crisp's findings show that candidates do use response space as an indicator of demand, it should also be noted that the display of the marks available also contributes to candidates' expectations of demand (Crisp, Sweiry, Ahmed & Pollitt, 2002). Combined booklets have other benefits including that they reduce the amount of paper candidates have to cope with (Crisp, 2008); may help candidates with time management and checking (Crisp, 2008); and allow for structured questions which can reduce the level of demand for less able candidates (Baird, Chamberlain, Meadows, Royal-Dawson & Taylor, 2009). For example, if the question asks the candidate to give two examples, the response space may be labelled 1 and 2 to reinforce the need for two examples. However, Crisp (2008) found that participants did report running out of space for long answer questions. The logical

solution to this problem is to simply provide more space for the response; yet providing more space may result in some candidates wasting time writing longer answers than required (Crisp *et al.*, 2002). Crisp (2008) concluded that combined booklets were preferable for many exam formats but separate booklets should be used for exams that require long answer questions, particularly for higher demand AS and A2 exams.

In the past, AQA have provided no guidance as to whether or not to use a combined format; traditionally GCSEs used combined booklets and A-levels used separate booklets. AQA now advise to use a combined booklet, if possible, for all exams. The ethos behind this advice is that combined formats provide an indication of the length of an appropriate response, as well as ensuring that all the information is in one place, i.e. preferable to separate documents and inserts. The secondary motivation for AQA to endorse combined booklets comes from the increased use of CMI+ marking. CMI+ marking requires combined booklet format and AQA plan on using CMI+ marking for almost all of their papers in the future. AQA have, however, encountered problems with using combined booklets. In the June 2009 series, A-level Geography was converted to a combined booklet format for CMI+ marking. On several of the questions, the candidates ran out of space and used extra pages, often answering two separate questions on the same extra page. The extra pages could not be easily identified, thus it was difficult and time consuming for AQA to marry up the extra pages with questions the candidates were answering. AQA need to ensure that this situation does not occur in future examination series. Currently, AQA are working on practical solutions to provide candidates with extra space.

3.3 Interaction between presentation and information processing

Information can be presented to candidates in a variety of forms: for example, text, diagrams, and tables can often be used to display the same information. The way in which information is presented can influence candidates' interpretation of the questions (Fisher-Hoch, Hughes & Bramley, 1997). Images have many benefits such as: making the abstract more concrete, motivating students to respond, and making the paper seem less daunting (Crisp & Sweiry, 2006; Johnson, 2004). On the other hand, the inclusion of images can lead to misinterpretation of the questions (Chamberlain, 2009). An attentional bias has been observed in highly anxious people, which results in them allocating attentional resources to particularly distracting information that may not be relevant to the task in hand (Fox, 1994). Images can be more salient to candidates than textual information and can therefore act as a distracter to candidates, should the image not be essential to answering the question (Crisp, Sweiry, Ahmed, & Pollit, 2008). Pollitt and Ahmed (2000) found that in instances where an image provided context but reference to it was not needed for a successful response, candidates lost marks by trying to make the image relevant. Crisp *et al*. (2008) provided a similar explanation; they argue that if an image is provided, candidates expect to use it in their answer and may misinterpret the question as a result. Crisp and Sweiry (2006) investigated the effect of using images in exam questions. They found that removing the image did not help candidates, instead it resulted in fewer candidates answering the questions. Interviews suggested that this was because the images led candidates to perceive the questions as easier. This could have been because the images facilitated the candidates' thinking. The images did not have a dramatic effect on misleading the candidates; contrary to Pollitt & Ahmed's (2000) findings, most seemed to work out that they did not directly relate to the answer.

Page layout can influence how candidates process information, meaning that the location of information on the page is important. Placing information higher on a page will make it appear

more important to candidates (Winn, 1987 cited in Crisp & Sweiry 2006). The first elements contained within a mental model will dominate and strongly influence subsequent elements, therefore information which is processed earlier in the paper can affect how candidates interpret information in questions which follow the information. Candidates expect the meaning from earlier information to be linked to information that follows; this can lead to candidates misinterpreting what is required of them (Johnson-Laird, 1983). Factors such as working memory and short term memory also affect candidates' performance. Research suggests that the information which is accessed first and last is most reliably recalled from the short term memory (Wiswede, Russeler & Munte, 2007; Murdock, 1962; Feigenbaum & Simon, 1962). With this in mind, important information may be best placed at the beginning and the actual task instruction at the end; however this review has found no literature which directly addresses this hypothesis. In the interviews that followed Crisp's (2008) research, the participants reported that the proximity of the response space to the questions helped them respond specifically to the question asked. Having the question close to where candidates write their answer reduces the need to remember the question, which may explain why candidates found this to be helpful. Once again, this finding advocates the use of a combined booklet format where the response space is necessarily nearer to the question.

An example of poor presentation is included in Appendix A. The information for this GCSE Geography question has been presented to the candidates in a confusing manner. Firstly, the instruction at the top of the page refers to two diagrams, one of which, Figure 2a, is two pages in front of the question in the paper, making simultaneous viewing impossible for the candidates. Secondly, Figure 2a does not add any more information than that already provided in the question. Candidates will expect the diagram to be providing necessary information that is additional to that stated in the question and therefore may be misled and include irrelevant information from Figure 2a. Thirdly, the task candidates are asked to perform is embedded in between factual information and the second diagram; the instructive text is not immediately apparent and may result in candidates being unsure of what they have to do. An alternative version of this question is suggested in Appendix B. This alternative version is likely to induce different responses from candidates simply by altering the page layout. Clearly, the presentation and sequencing of information needs to be considered when producing exam papers.

The research into page layout also has implications for online assessment, where A4 pages cannot be viewed in their entirety and candidates are required to scroll down the page. Spatial location of information on the page has been found to act as a prompt for memory in word recall tests (Lovelace & Southall, 1983). Wastlund, Norlander and Archer (2008) adapted a paper based assessment so that only the presentation of the paper (on-screen versus printed paper) varied between two conditions. They found that candidates taking online examinations had an increased workload compared to paper based candidates. Wastlund *et al*. (2008) speculated that this was due to the on-screen candidates being deprived of spatial clues. Whilst this study cannot decisively determine the cognitive processes which underlie their findings, the conclusions drawn are in keeping with Lovelace and Southall's (1983) research.

3.4 Summary of presentation effects

Research suggests that the font, formatting and layout have an impact on the readability of exam papers. Size 12 Arial font is preferred by children and there is some evidence to suggest that it is better for reading comprehension than other typefaces. There is evidence to suggest that the space provided for a response is used by candidates as an indication of how detailed the examiner intends their response to be. Combined booklets have been seen to induce fuller,

more focused responses from candidates than exam papers with separate answer sheets. The medium in which information is presented has been seen to interact with candidates' expectations and information processing, which subsequently affect how they answer. If candidates expect to be asked questions in a certain way, for example using a diagram which is provided, then they may use it in their response regardless of whether it is relevant to the question. However, this does not occur in all cases and there are benefits of using images. Sequencing of information also affects how candidates process questions and respond to them. The effect of the above presentational features does depend upon the given question and situation. Whilst the research reviewed thus far can be taken as a general guide, hard and fast rules cannot be applied to paper writing. The general advice which can be drawn from the literature reviewed, together with AQA current practices and future actions, are summarised in the table below.

Research suggests	AQA practice	AQA action
Font should be clear; Arial size 12 is advised, although the difference in readability between size 11 and 12 is negligible.	By June 2010, AQA's papers will be printed in Arial size 11. Size 11 was chosen over 12 due to concerns of exam paper length.	N.A.
Combined booklets are preferable to separate answer booklets, although in some cases (for example long answer questions), separate booklets are preferable.	Currently, AQA advises the use of combined booklets, particularly in light of the move to CMI+. Some difficulties with using combined booklets have arisen. Solutions for providing extra answer space are being considered.	AQA is continually reviewing and enhancing CMI+ marking.
Layout of information needs to be considered: placing key information at the top of the page; ensuring the instructions for the task are the last thing candidates read before they respond; taking into account the effect of preceding questions.	The layout of information and the use of stimuli differ across AQA papers. Whilst no set procedures can be enforced due to the varying nature of papers, more guidance could be provided for question paper writers. However, most AQA Principal Examiners have naturally adopted these rules for writing papers and, as a result, most AQA papers are currently written with consideration to information layout and appropriate use of stimuli.	More guidance to be provided for question paper writers regarding layout of information.

4. STRUCTURE

There are a variety of considerations which surround question paper structure: question order, mix of question types, varying format, and optional questions (to name a few). Question paper structure is a somewhat forgotten area in today's research. However, there was a great deal of interest in paper structure in the 1970s and 1980s; research which is still relevant today.

4.1 Question types

Performance on an exam does not only reflect knowledge in a given subject but also the ability to answer questions (Murphy, 1980). Candidates' performance on a paper will, to some extent,

depend upon the types of questions asked. Questions come in a variety of forms: for example, short answer, essay based, and multiple-choice. Gustav (1964) investigated undergraduate psychology students' question type preferences. The participants reported preferring exam papers where essay questions were used over multiple-choice as they felt they provided more opportunity to display their knowledge. The participants also reported being more motivated to study for essay based exams than true or false or multiple-choice formats. It is worth noting that undergraduate psychology students are likely to have different preferences to GCSE and GCE candidates taking different subjects. Interestingly, Gustav found few differences in the participants' actual performance on essay based and multiple-choice formats, despite their reported preference and motivation for essay based exams. Candidates' differ in learning style, which is determined by the individual characteristics of the candidate, their motivation for study, and the type of assessment. Some candidates are inclined to take a surface approach to learning, covering the required information in order to achieve the desired grade, whilst other candidates take a deep approach, learning for the sake of learning in order to understand a given topic and synthesise their knowledge. Learning style will influence preferences for exam format; students who tend towards a surface strategy will also have a tendency to prefer multiple-choice format whilst deep learners tend to prefer essay based exams. Equally, exam format will impact upon learning style; deep learners may take a surface strategy to studying for a multiple choice exam and, conversely, surface learners may switch to a deep learning strategy for an essay based exam (Daly & Pinot de Moira, 2008).

The type of questions included in an exam paper will, to some extent, be dictated by the skills that are being assessed (Stringer, 2009). Essay based questions are historically thought to be better at assessing divergent skills whilst multiple-choice questions are thought to be better at assessing breadth of knowledge in a subject (O'Donovan, 2005; Curren, 2004). If a paper is constructed entirely from one type of question then there is a danger of it including a systematic bias. Past research has demonstrated sex differences for some question types. Males have been observed to do better than females in multiple-choice objective exams whilst females do better than males in open-ended exams (Murphy, 1980). Wester and Henriksson (2000) found that males and females did not differ in performance on two versions of a maths test; one version being multiple-choice, the other being open-ended short answer questions. They concluded that no result was found because the short answer questions and multiple-choice questions were assessing the same skills. It has been theorised that sex differences for multiple-choice questions are due to the different cognitive styles of males and females; males are better at convergent thinking whilst females are better at divergent thinking (Murphy, 1980). In Wester and Henriksson's (2000) study, the questions in both conditions may have been assessing convergent thinking which is associated with maths. In some cases, therefore, gender bias may be legitimate; however, it is important that exam results reflect candidates' abilities at a given subject and are not the result of biases that are not related to the subject. The area of sex bias in multiple-choice questions is highly controversial and it should be noted that there is a body of research which contests the sex differences. Equally, multiple-choice questions have evolved to test a variety of skills and now come in many forms (Stringer, 2009; Curren, 2004). Nevertheless, the simplest method of avoiding systematic bias is to include a mix of question types in the paper, where appropriate.

Optional questions were introduced in acknowledgment of the fact that different questions suited different candidates. The rationale behind the introduction of optional questions was to facilitate educational gain through allowing teachers to teach their specialties, and candidates to pick questions which best showed their ability (Willmott & Hall, 1975). Indeed, Johnson's (2004) research into optional questions found that candidates liked question choice; they reported that

they liked to be able to pick the question that suited them. More recently, Eason (2006) conducted a survey with centres which offered GCE Geography to determine which factors influenced their decision to choose an awarding body. The feedback from the survey indicated that the option of choice was highly influential to their decision making; centres like specifications which offer a choice of topics.

There are, however, drawbacks to using optional questions. It is almost impossible to write optional questions of precisely the same difficulty and this can lead to problems for comparability. Taylor's (2009) research illustrates how hard it is to write comparable optional questions. A Rasch analysis was conducted on a GCSE English paper containing optional questions. The questions were carefully selected so as to assess the same skills and to be of equal difficulty. The analysis revealed that whilst the examiners had ensured that the optional questions were of equal difficulty in terms of communication and organisation, the difficulty of the questions differed in terms of sentence structure, punctuation and spelling. This was not expected given that the marks are awarded for the same construct using the same criteria. This example demonstrates that the difficulty of optional questions is unpredictable even to experienced examiners (Taylor, 2009). Optional questions were introduced on the assumption that candidates will be able to select the questions which they are best equipped to answer. However, research has shown that candidates attempt questions sequentially; often selecting the first question that they can answer rather than assessing all of the options (Wilmut, 1980a). Research into planned behaviour suggests that this could be due to candidates having too much choice. In situations where people are faced with a high level of choice, they are hindered from making optimal choice decisions; they struggle to cope with the amount of information and therefore are less likely to make fully informed decisions (Ajzen, & Fishbein, 1980 cited in Parkinson, 2009). Equally, several studies have shown that more able candidates are slightly better at picking easier questions than less able candidates (Willmott & Hall, 1975; Francis, 1977a; Wilmut, 1980a). Still, in cases where optional questions are used, question choice appears to affect outcome. Willmott and Hall (1975) demonstrated that the final scores on a biology examination could vary by as much as 5 to 7.5 *per cent* of the total marks due simply to question selection. Candidates taking exams with optional questions also encounter the added stress of having to choose between them (Baird *et al*., 2009; Francis, 1977b). Whilst optional questions are desired by centres and candidates alike, careful consideration needs to be given as to how to fairly award candidates' grades.

4.2 Order of questions

The structure of a question paper will be influenced by the order of the questions. It may be beneficial to order questions in terms of difficulty. There is a fairly constant finding in the literature that an easy-to-hard question arrangement results in better performance than hard-to-easy question arrangement (Leary & Dorans, 1985). There are several reasons why ordering questions easy-to-hard benefits candidates. Firstly, anxiety may play a role; difficult questions at the beginning of the exam could heighten anxiety levels and, as previously mentioned, high levels of anxiety reduce exam performance (Mulkey & O'Neil, 1999). Placing easy questions at the beginning may reduce anxiety and enhance exam performance. However, there have been mixed findings regarding ordering questions by level of demand and the affect this has on anxiety levels. Some authors may not have found an effect due to the anxiety measures used for such investigations. Anxiety measures are designed for use in everyday situations and do not accurately measure candidates when they are under unusually high levels of anxiety, i.e. exam conditions, which could account for why some authors have found no effect (Tippets, 1989). Secondly, easy-to-hard arrangement may improve performance as candidates'

perceptions of their performance on earlier questions can affect their motivation and performance on later questions; if they struggle on harder questions early on then they may be deterred from trying questions later in the paper. Thirdly, candidates typically perform worse at the end of exams, most likely due to fatigue and poor time management (Willmott & Hall, 1975; Wilmut, 1980a). As previously mentioned, candidates answer questions sequentially; therefore if easier questions are placed towards the end of a paper, poorer candidates may either not get to them or perform badly on them, thus missing out on marks for questions that they could have answered if they were positioned nearer the start of the paper (Wilmut, 1980a). Several studies involving tests with no time constraints have found that the fall in performance at the end of the examination disappears which indicates that time factors have a strong negative impact on performance at the end of the examinations (Leary & Dorans, 1985). Time pressures are not a factor for high stakes examinations where ample time is allowed for completion. A recent investigation into a GCSE Science multiple-choice paper showed no evidence of time pressures negatively affecting candidates. The paper was arranged easy-to-hard and whilst non-completion rates moderately increased towards the end of the paper, due to the questions getting harder, the guessing index demonstrated no evidence for the random or careless selection of answers (Stringer, 2009).

An alternative approach to paper structure is to order questions by topic or theme. Commonly, examination paper structure reflects the specification structure, with sections of the paper covering topical areas. This arrangement reflects candidates' learning; candidates, particularly at GCSE level, compartmentalise their learning and are not expected to think synoptically. The drawback to this approach is that candidates may not perform as well on topics that are at the end of the paper, which may disadvantage candidates whose strengths lie in those specific topics (Wilmut, 1980a). Sax and Carr (1962) investigated the effect of organising questions by subject versus difficulty. Participants attempted significantly more questions and obtained higher scores when questions were ordered by increasing difficulty than when they were arranged by topic. However, the researchers used a test (the Henman-Nelson Test of Mental Ability) that is designed to measure those aspects of mental ability that are important for success in academic work and in similar endeavours outside the classroom. As such, the participants would not have learnt the content in a structured manner. Therefore, it may be that participants did not benefit from the exam being structured by subject as this did not reflect how they acquired the knowledge.

Some exams are structured by linking questions together according to themes in an attempt to facilitate candidates' recall by building on answers to the previous questions. Typically, structured question format is used to link questions together, with question sub-parts being linked to a central theme stated in the stem (Wilmut, 1980b). Fisher-Hoch *et al*. (1997) found that altering the sequence of mathematics sub-questions made the questions more difficult. They concluded that the order in which a problem occurs is important and the original sequence was the most appropriate way to order the questions. Clearly, in Fisher-Hoch *et al*.'s study the original sequence of the question sub-parts facilitated candidates' responses. Conversely, other research suggests that linking questions together by a theme can impede candidates' performance. Linking by theme has been found to confuse candidates if the linking results in questions being asked in an unfamiliar context. Crisp *et al*. (2002) found that the context of previous questions, which were on biology, confused candidates when the follow-on questions related to a different topic, in this case physics. Such confusion can lead to candidates misinterpreting questions (Frazier & Rayner, 1982). The appropriateness of linking questions may be determined by the subject. Linking questions may also deter candidates from attempting later questions if they have found the preceding questions difficult. Candidates see the

questions as connected if they are on the same topic and, as a result, if they have struggled on an earlier question on the same theme, their performance is affected on the later question (Wilmut, 1980b).

For certain types of exams, it may be appropriate to organise the paper by task types. Murphy (1978) investigated the effect of changing the sequence of a French oral exam. The French oral exam involved four sections: three where candidates had to respond to questions based on unseen material, and the fourth where candidates participated in a role play which they were given to study fifteen minutes before the start of the examination. Murphy found that candidates did better if the pre-prepared work was put at the beginning. He hypothesised that candidates performed worse when this section appeared later as they were distracted by trying to remember the role-play scenario. This is in keeping with cognitive research into prospective memory. Researchers have found that participants perform worse on cognitive tasks if they are carrying a 'prospective memory load', trying to remember to perform a task in the future (Smith, 2003). Murphy's (1978) work may extend to written exams which involve pre-prepared components, for example the first question on the AQA GCSE English B paper requires candidates to answer questions using a pre-release booklet (see Appendix C). Positioning the pre-release material at the beginning of the exam could ensure that performance on questions two and three are not adversely affected by candidates trying to remember key points concerning the pre-release material.

4.3. Predictability

Some specifications maintain a fixed structure to their exams year on year, where similar questions to earlier years appear in the same predictable order as previous papers; frequently this is the case for exam papers which are organised by task type. Candidates often spend a significant amount of time preparing for examinations, using past papers to practise, and being instructed on the best ways to respond (Baird *et al.*, 2009; Broekkamp, Van Hout-Wolters, Van den Bergh, & Rijlaarsdam, 2004). The dominance of past papers in candidates' revision strategies means that they often expect the questions in the exam to be like those of previous years (Baird *et al.,* 2009). Candidates develop schema: a pre-defined framework of what topics the paper will include, the way in which the questions will be asked, and the kinds of responses that will be required. Exam papers which maintain a fixed structure are likely to encourage this type of examination preparation and activate candidates' exam schemas during the exam. The more familiar the style of the paper, the stronger a schema is evoked (Crisp *et al.*, 2002). These expectations can help candidates to reduce anxiety and uncertainty and successfully navigate the paper; however, expectations can become problematic if the questions do not fit the candidate's schema. In such cases, the schema can interfere with the processing of the question, resulting in candidates answering the questions they were expecting to find rather than the questions which are actually in front of them (Crisp *et al.*, 2008). Baird *et al*. (2009) provided an alternative explanation for candidates inappropriately responding to questions. Interviews were conducted with post A-level exam candidates who reported preparing answers as part of their exam preparation. Baird *et al*. (2009) concluded that some candidates prepared their responses to such a degree that they were unable to respond to anything unexpected. In such cases, the continuation of exam predictability is encouraging examination preparation techniques that are inflexible. Whilst the evidence discussed so far refers to the individual questions, the findings might extend to the overall paper structure; papers which follow the same structure year on year are likely to encourage the exam preparation discussed in Baird et al (2009). The negative impact of predictability needs to be weighed up against the negative impact on candidates if the paper structure were to change dramatically year on year. If the

paper structure were to change suddenly then a form of construct-irrelevant variance may be introduced. The exam will no longer be just assessing candidates' ability in a given subject, it will also be assessing candidates' ability to respond to a novel exam format. An element of predictability in exam structures is therefore desirable and a careful balance must be maintained (Baird *et al*. 2009).

4.4. Summary of structure

The type of questions included in the paper and the order in which they are presented are the chief concerns when structuring papers. The types of question used in a paper will be primarily based upon the skills and knowledge that the paper is intended to assess. However, the literature demonstrates that question types affect candidate performance; different candidates perform best on different question types. Thus, where appropriate, a mix of question types is preferable to provide candidates with different opportunities to demonstrate their knowledge. Equally, as candidates prefer different types of question, they also like to be given a choice of questions. Research cautions against the use of optional questions as candidates can be disadvantaged by their question choice. Poorer candidates do not always select the best question to display their knowledge which results in a significant drop in performance that is not related to their ability in the given subject. Careful consideration needs to be given before using optional questions in an exam paper; discussing their desirability with centres would be worthwhile. The majority of the literature available on question order relates to arranging questions by difficulty. The literature suggests that an incline in difficulty results in better performance from candidates. Commonly, exam structure reflects the specification and questions are ordered by topic. There is tentative evidence to suggest that ordering by topic may disadvantage candidates if it results in questions not following an incline in difficulty; however, there is very little research which has directly addressed ordering by topic. Similarly, there is evidence which tentatively suggests that questions relating to pre-prepared work should be positioned near the beginning of the paper in order to improve candidates' performance. How closely the structure of the paper relates to previous years needs to be carefully considered. Maintaining a constant structure can benefit candidates through reducing anxiety and making the paper easier to navigate. On the other hand, a constant structure may encourage strong expectations from candidates regarding the exam content and promote prepared inflexible responses which may not relate to the actual question. Again, whilst the research reviewed thus far can be taken as a general guide, hard and fast rules cannot be applied to paper writing. The general advice which can be drawn from the literature reviewed, together with AQA current practices and future actions, are summarised in the table below.

Research suggests	AQA practice	AQA action
The types of question included in the paper should assess the required skills and knowledge, and vary in style/format.	Currently, AQA does not provide question paper writers with procedural guidance for structuring papers. However, unwritten rules of good practice have evolved over the years and most of AQA's papers do naturally follow the guidelines outlined by the research literature. AQA plan on altering their procedures for question paper writing. Examiners may be restricted by these new processes and unable to continue to practise the unwritten rules of paper writing they have developed over the years.	It would perhaps be worthwhile for AQA to make these unwritten rules explicit to ensure a more standardised practice across subject departments, as well as assisting inexperienced paper writers. Equally, AQA need to have a good understanding of how their papers will be affected by the new process of question paper writing.
Question papers should have an incline of difficulty.		
Questions which relate to pre-prepared work should be positioned near the beginning of the paper.		
Exams should be structured in a familiar style to candidates whilst not being overly predictable.		
Optional questions are attractive to teachers and students but comparability is nigh impossible to ensure.	Optional questions are frequently used in AQA's papers.	It may be beneficial for AQA to investigate new practices, for example post-hoc scaling methods to improve comparability.

5. CONCLUSIONS

This review provides a broad overview of a wide range of presentational and structural issues. There is very little current literature on question paper presentation and structure, and the literature from the 1970s-1980s is scarce. Many structural and presentational considerations have not been discussed as, to date, they have not been fully investigated; the effect of the format of numbering in questions and subject specific considerations, for example, have not been discussed. As a result, the conclusions which can be drawn from this literature review are tentative.

The research suggests that presentational and structural considerations affect candidates' performance. If candidates misread questions, misinterpret the requirements of the given task, or miss questions which they are capable of answering due to either presentation or placement in the paper, then the paper is no longer accurately assessing ability. If the types of questions in the paper favour certain candidates over others then the paper is also assessing the candidates' ability at answering a given question type. If candidates' performance is affected by which optional questions they choose to answer, then the paper is also assessing the candidates' ability to select suitable questions. If the paper is no longer solely assessing what it is intended to assess, then it loses validity. An examination paper is not simply as good as the questions within it; presentation and structure make it greater than the sum of its parts.

AQA's current practices were discussed in light of the general guide generated from this research, revealing potential areas which AQA may choose to address. Two areas of possible further research were identified: refinement on CMI+ marking and optional question comparability. Firstly, the research literature and AQA's own experience testify to the fact that a combined booklet format is problematic for some papers; issues arise when candidates require

different amounts of space for their answers and extra sheets are needed. Further development of CMI+ marking could be investigated in order to produce practical solutions when extra sheets are used. Secondly, the research literature informs us that whilst optional questions are desired by centres and bring educational benefits to candidates, comparability cannot be ensured when they are used. It may be beneficial for AQA to investigate new practices, for example post-hoc scaling methods to improve comparability.

In light of this review, what work does AQA need to do before implementing question banking for their papers? Firstly, it may be worthwhile for AQA to make current unwritten rules of question paper writing explicit to ensure a more standardised practice across subject departments and to ensure the paper quality is not lost. This literature review has only touched upon some of the issues surrounding question paper presentation and structure. The conclusions which can be drawn are somewhat limited. More research is required to effectively inform question paper writers. Questions need to be raised as to whether a 'one-size fits all' philosophy is appropriate for question paper construction. The literature reviewed so far has found varying effects of question paper presentation depending upon question types in the exam. Equally, discrepancies between studies investigating the effect of question paper structure can be explained by different test content. Would the advice for a science paper differ to that for a history paper? More needs to be known about content specific effects. Secondly, it would be beneficial for AQA to have a good understanding of how their papers will be affected by the new process of question paper writing. Will a question banking process affect how question papers are structured? Whitehouse *et al*. (2008) found that whilst on the whole examiners felt that there were considerable benefits to using question banking test construction, they did feel limited as they could not alter the questions to fit their paper. It is possible that question banking test construction would result in less coherent question papers.

Taking into account the gaps in the empirical literature, and the changing processes within AQA, this paper accordingly proposes that research be conducted into good practice for question paper structure. AQA will need to access the current good practice and processes for question paper construction for a variety of paper types and investigate how question banking test construction will alter these processes. The findings from this research will provide evidence to guide and inform educational researchers, policy makers, and curriculum and assessment developers.

Victoria Spalding
19.09.09

REFERENCES

Baird, J., Chamberlain, S., Meadows, M., Royal-Dawson, L., & Taylor, R. (2009). Students' views of stretch and challenge in A-level examinations. Guildford: Assessment and Qualifications Alliance.

Bernard, M. L., Chaparro, B. S., Mills, M. M., & Halcomb, C. G. (2003). Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. *International Journal of Human-Computer Studies, 59*, 823-835.

Bernard, M. L., Chaparro, B. S., Mills, M. M., & Halcomb, C. G. (2002). Examining children's reading performance and preference for different computer-displayed text. *Behaviour and Information Technology, 21*(2), 87-96.

Broekkamp, H., van Hout-Wolters, B. H., van den Bergh, H., & Rijlaarsdam, G. (2004). Teachers' task demands, students' test expectations, and actual test content. *British Journal of Educational Psychology 74*(2), 205-20.

Case, S., & Swanson, D. (2001). *Constructing written test questions for the basic and clinical sciences.* Philadelphia: National Board of Medical Examiners.

Chamberlain, S. (2009). A guide to writing examination questions: theoretical approaches and practical solutions. Guildford: Assessment and Qualifications Alliance.

Crisp, V. (2008) Improving students' capacity to show their knowledge, understanding and skills in exams by using combined question and answer papers. *Research Papers in Education, 23*(1), 69-84.

Crisp, V. & Sweiry, E. (2006) Can a picture ruin a thousand words? The effects of visual resources in exam questions. *Educational Research 28*(2), 139-154.

Crisp, V., Sweiry, E., Ahmed, A., & Pollitt, A. (2002) *Tales of the expected: the influence of students' expectations on question validity and implications for writing exam questions.* Paper presented at the British Educational Research Association conference, Exeter.

Crisp, V., Sweiry, E., Ahmed, A., & Pollitt, A. (2008). Tales of the expected: the influence of students' expectations on question validity and implications for writing exam questions. *Educational Research, 50*(1), 95-115.

Curren, R. R. (2004). Educational measurement and knowledge of other minds. *Theory and Research in Education, 2*(3), 235-253.

Daly, A., Pinot de Moira, A. (2008). Students' approaches to learning and their performance in the Extended Project Pilot. Guildford: Assessment and Qualifications Alliance.

Eason, S. (2006). GCE Geography specification development questionnaire survey, analysis of responses to closed questions. Guildford: Assessment and Qualifications Alliance.

Feigenbaum, E. A., & Simon, H. A. (1962). A theory of the serial position effect. *British Journal of Psychology, 53*(3), 307-320.

Fisher-Hoch, H., Hughes, S., & Bramley, T. (1997, September). *What makes GCSE exam questions difficult?: Outcomes of manipulating difficulty of GCSE questions.* Paper presented at the British Educational Research Association annual conference, York.

Fox, E. (1994). Attentional bias in anxiety: A defective inhibition hypothesis. *Cognition and Emotion*, *8*(2), 165-195.

Francis, J. (1977a). An investigation into question choice in ordinary level chemistry. Guildford: The Associated Examining Board.

Francis, J. (1977b). A further investigation into question choice. Guildford: The Associated Examining Board.

Frazier, L., & Rayner, K. (1982). Making and correcting errors in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178-210.

Green, S. (2003). Exploring a paradox in children's writing: An investigation of evidence which suggests that task support does not have the desired effect. *Education, 31*, 19-21.

Gustav, A. (1964). Students' preferences for test format in relation to their test scores. *The Journal of Psychology, 57*, 159-164.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of mulitiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309-334.

Hartley, J., & Trueman, M. (1986). The effects of the typographic layout of close-type test on reading comprehension scores. *Journal of Research in Reading 9*(2), 116-24.

Huck, S. W., & Jacko, E. (1974). Effect of varying the response format of the Alpert-Haber Achievement Anxiety Test. *Journal of Counselling Psychology, 2* (2), 159-163.

Hughes, S., Pollitt, A., & Ahmed, A. (1998). The development of a tool for gauging the demands of GCSE and A level exam questions. *BERA.*

Johnson, M. (2004) This one is more me! What children think about writing test stimuli involving choice. *Literacy, 38*(2), 90-96.

Johnson-Laird, P.N. (1983). Mental models: Towards a cognitive science of language, inference and consciousness. Cambridge, MA: Harvard University Press.

Knupfer, N. N., & McIsaac, M. S. (1992). Designing instructional materials with desktop publishing software: The effect of white-space variation on learning. *Journal of Research on Computing in Education 28*(1), 75-87.

Leary, L. F., & Dorans, N. J. (1985). Implications for altering the context in which items appear: A historical perspective on an immediate concern. *Review of Educational Research, 55*, 387-413.

Lovelace, E. A., Southall, S. D. (1983). Memory for words in prose and their locations on the page. *Memory and Cognition, 11*(5), 429-434.

Mulkey, J. R., O'Neil, H. F. Jr. (1999). The effect of test item format on self-efficacy and worry during a high-stakes computer-based certification examination. *Computers in Human Behaviour, 15*, 495-509.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482-488.

Murphy, R. (1978). Report of an investigation into the effect of the order of the parts of the oral examinations in French, German, and Italian at the ordinary level. Guildford: The Associated Examining Board.

Murphy, R. (1980). Sex differences in objective test performance. Guildford: The Associated Examining Board.

O' Donovan, N. (2005) There are no wrong answers: an investigation into the assessment of candidates' responses to essay-based examinations. *Oxford Review of Education, 31*(3), 395-422.

Parkinson, S. (2009). The theory of planned behaviour. Guildford: Assessment and Qualifications Alliance.

Paterson, D. E., & Tinker, M. A. (1932). Studies of typographical factors influencing speed of reading. *Journal of Applied Psychology*, *XVI*, 605-613.

Pollitt, A., & Ahmed, A. (2001, May). *Science or Reading?: How students think when answering TIMSS questions*. Paper presented to the International Association for Educational Assessment, Rio de Janeiro, Brazil.

Pollitt, A., & Ahmed, A. (1999, May). *A new model of the question answering process.* Paper presented to the International Association for Educational Assessment, Bled, Slovenia.

Pollitt, A., & Ahmed, A. (2000, September). *Comprehension failures in educational assessment.* Paper presented at the European Conference on Educational Research, Edinburgh.

Poulton, E. C. (1965). Letter differentiation and rate of comprehension of reading. *Journal of Applied Psychology*, *49*, 358-362.

Sanocki, T. (1991). Looking for a structural network: Effects of changing size and style on letter recognition. *Perception 20,* 529-541.

Sax, G., & Carr, A. (1962). An investigation of response sets on altered parallel forms. *Educational and Psychological Measurement, 22*, 371-376.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 347-361.

Stringer, N. (2009). Response to Ofqual science report. Guildford: Assessment and Qualifications Alliance.

Taylor, R. (2009). GCSE English A (3702): Some insight gained from the Rasch model. Guildford: Assessment and Qualifications Alliance.

Tippets, E. (1989). The effect of item arrangement on test anxiety. *Applied Measurement in Education, 2*(4), 289-296.

Wastlund, E., Norlander, T., Archer, T. (2008). The effect of page layout on mental workload: A dual-task experiment. *Computers in Human Behaviour, 24*, 1229-1245.

Wester, A., & Henriksson, W. (2000). The interaction between item format and gender differences in mathematics performance based on TIMSS data. *Studies in Educational Evaluation, 26*, 79-90.

Wheadon, C. B., Whitehouse, C., Spalding, V., Tremain, K., & Charman, M. (2009). Principles and practice of on-demand testing. Guildford: Assessment and Qualifications Alliance.

Whitehouse, C., He, C., & Wheadon, C. B. (2008). Item banking with a test construction interface: An evaluation of a prototype. Guildford: Assessment and Qualifications Alliance.

Willmott, A. S., & Hall, C. G. W. (1975). O-level examined: The effect of question choice. Schools Council Research Studies. Macmillan Education.

Wilmut, J. (1980a). Candidates' question choice in the June 1980 Religious Studies syllabus II (Multi-faith), Paper 1. Guildford: The Associated Examining Board.

Wilmut, J. (1980b). The use of structured questions in GCE examinations. Guildford: The Associated Examining Board.

Wiswede, D., Russeler, J., & Munte, T. F. (2007). Serial position effects in free memory recall—An ERP-study. *Biological Psychology, 75*, 185-193.

APPENDIX A:

Figure 2a, Question 2, Page 8 from Edexcel GCSE Geography Spec A June 2007

(b) Study Figure 2a. It is a map showing the temperature, rainfall and location of a number of settlements in Europe.

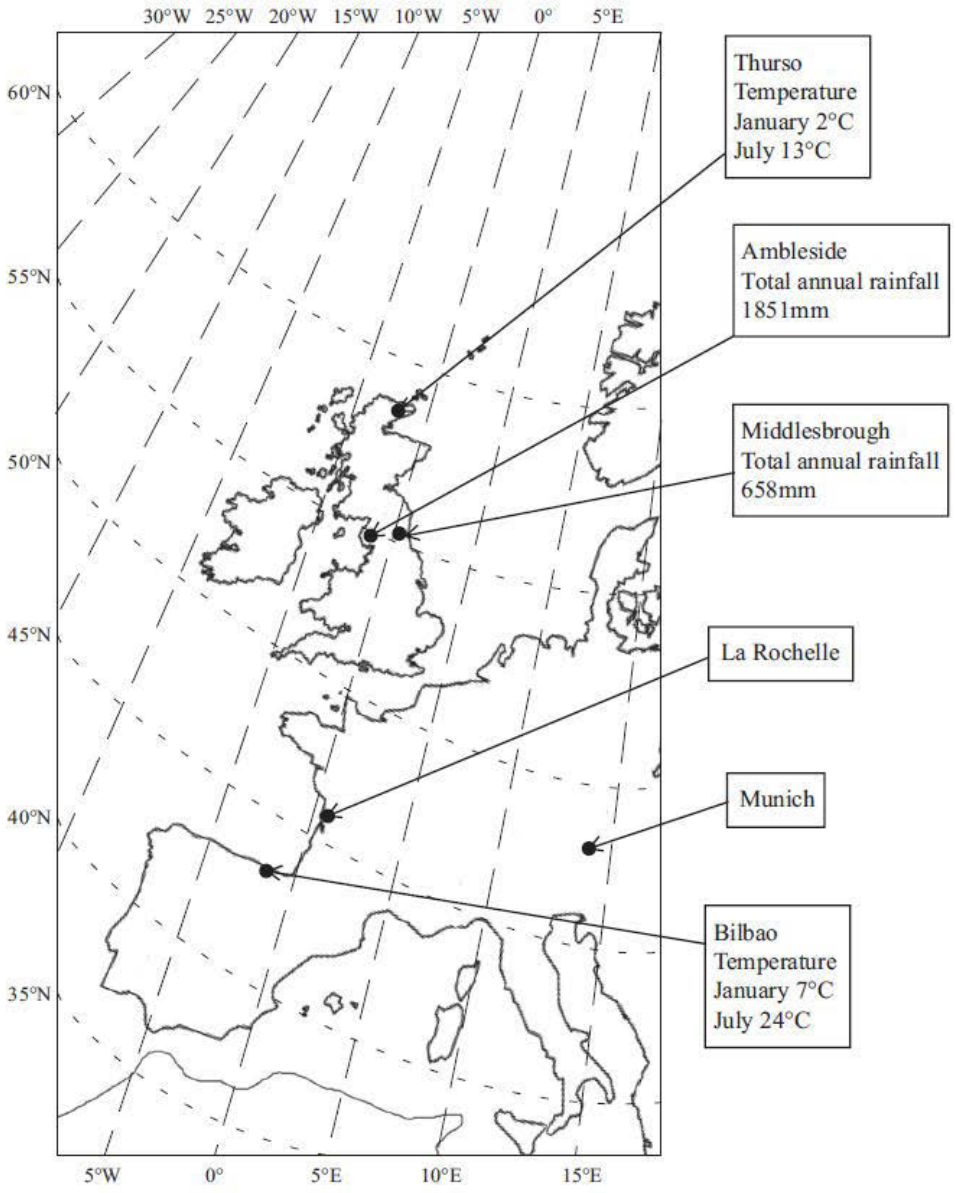

Figure 2a

Question 2b iv, Page 10 from Edexcel GCSE Geography Spec A June 2007.

(iv) Look again at Figure 2a. Also look at Figure 2b, a cross section of Northern England.

Ambleside receives more rainfall than Middlesbrough.

Use an annotated diagram or diagrams to explain why.

Use the space below Figure 2b for your answer

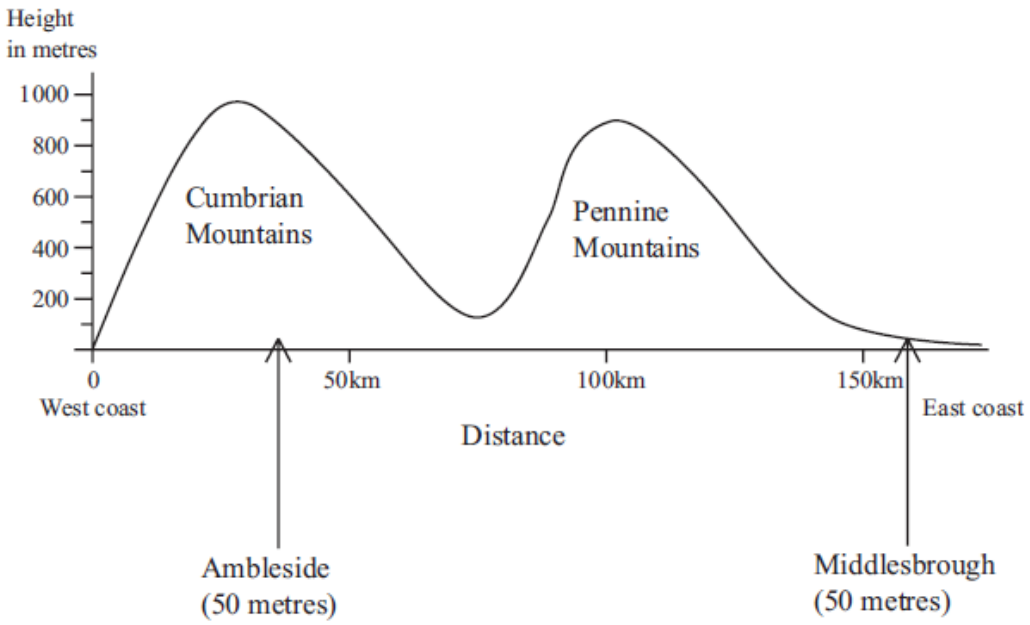

Figure 2b

(iv) Point mark 4

A clearly separate paragraph either above or below the diagrams is not an annotation and should not be credited. If statements are within the body of the diagrams without an appropriate arrow they should be credited.

2 mark max. on descriptive comments.

Expect explanation of relief rainfall. Air rises over the hills and cools. (1)

Condensation occurs. It rains.(1) This occurs before the clouds reach Middlesbrough therefore Ambleside is wetter. (1)

Annotations / labels to be on diagram do not accept separate paragraph.

APPENDIX B:
An amended question from Edexcel GCSE Geography Spec A June 2007

(iv) Look at Figure 2b, a cross section of Northern England.

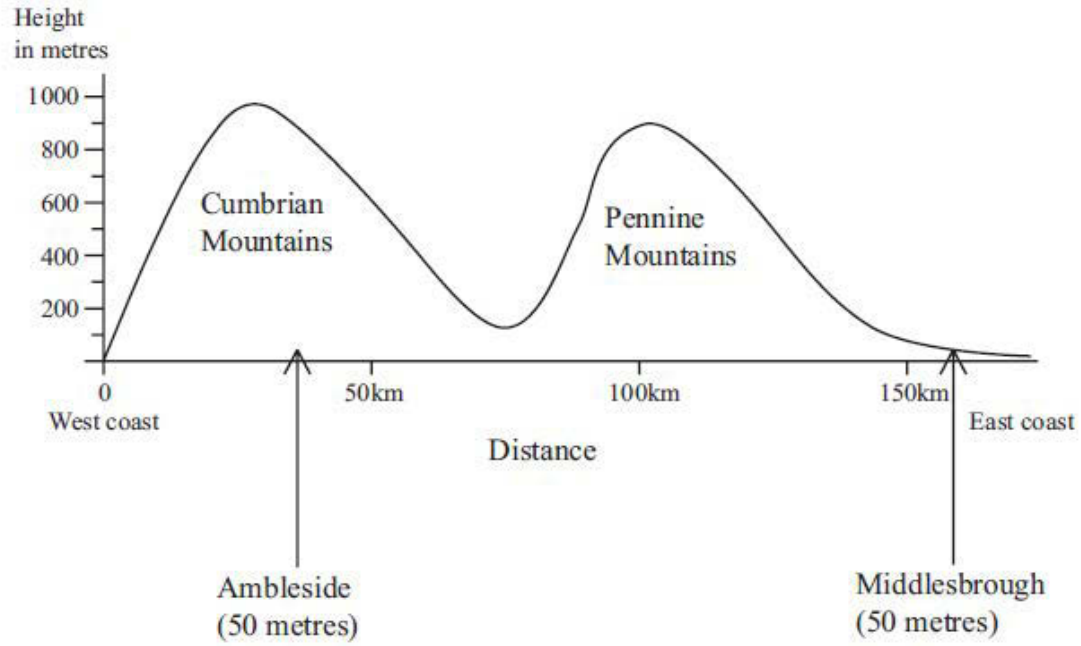

Figure 2b

Ambleside receives more rainfall than Middlesbrough.

Use an annotated diagram or diagrams to explain why.

APPENDIX C:
AQA GCSE English B June 2008

2

SECTION A: READING

MEDIA AND NON-FICTION TEXTS

Answer **both** questions in this section.

You are advised to spend about one hour on this section.

This includes 10 minutes reminding yourself of the content of the pre-release booklet.

You will be assessed on the quality of your Reading.

1 Media Texts

In your pre-release booklet look again at 'I Live With 70 Other Widows' on page 5 and at the text dealing with 'The police cell' and 'The prison cell' on page 7.

What differences do you find in these texts?

Write about:

- the content of each text
- the use of pictures
- fact and opinion in the texts. *(13 marks)*

2 Non-fiction Text

Read the extract on the page opposite, which comes from Hugo Hamilton's book, *The Speckled People*, in which the writer recalls features of his home life.

What differing approaches to home life do Hamilton's mother and father display? How does the language and structure of the extract bring out the differences? *(14 marks)*

4

SECTION B: WRITING TO ARGUE, PERSUADE, ADVISE

You are advised to spend about 40 minutes on this section.

You will be assessed on the quality of your Writing.

3 It is sometimes said that young people now remain at home for too long and fail to establish an independent life.

Write a newspaper article in which you argue for or against the view that young people have the right to remain at home for as long as they wish. *(27 marks)*