Date of meeting	3 June 2010
Agenda no.	10
Paper no.	RPA_10_CW_RP_012
Title	Reliability of on-screen marking of essays
Author(s)	Claire Whitehouse, Senior Research Associate

Background

This paper reports an opportunistic investigation of the on-screen marking reliability of a unit in AS Level Geography that contains a number of items requiring essay responses. The need for this investigation springs from the operational drive to migrate the marking of all scripts to on-screen by the summer of 2012. To help meet this target the e-marking application was modified to support the marking of essays on screen. The modifications were based on the outcomes from previous research projects. A small number of A Level units were part of a pilot of the modified application in January 2010.

The research used a group hierarchical true mark as the basis for the calculation of reliability indices. A small number of senior examiners formed the hierarchical group which re-marked items from 173 scripts on-screen. The segmented nature of the re-marking exercise was a better match for the reality of operational on-screen marking.

The report concludes that the mark/re-mark reliability of on-screen marking of essays and the question paper containing them is adequate and comparable with previously published values. Given the lack of a fully designed study, however, caution should be used when attempting to generalise from this set of results.

Included in the target audience are those responsible for the expansion of on-screen marking within AQA, AQA's technology partner for e-marking, subject managers and examiners. The paper is presented in its final version, subject to comments from the Research Committee.

Discussion points

- Given the expense and artificial nature of experimental studies of on-screen marking reliability, is the methodology used here sufficiently robust to provide convincing data to support the continued roll out of on-screen marking?
- Is the use of a group hierarchical 'true' mark a sensible approach in determining on-screen marking reliability? If not, then what is?
- Is it valid to use the conventional reliability indices of correlation coefficient, absolute mark difference and level of agreement for segmented, on-screen marking?
- Is the validity of marking more relevant than reliability? Does validity of marking in the on-screen mode always have to be compared with validity in the paper-based mode?

RELIABILITY OF ON-SCREEN MARKING OF ESSAYS

ABSTRACT

This paper reports the results of a re-marking exercise designed to explore the reliability of on-screen marking of a question paper that contains essay-type responses. As part of the study a new definition of the 'true' mark was proposed for the reporting of reliability indices. The new definition is better suited to segmented marking. It defines the 'true' mark for a script as the total mark based on marking contributions from a group of senior examiners, giving a reliability based on a group hierarchical 'true' mark. The hierarchical group in this study consisted of six senior examiners who re-marked a sample of 173 AS level Geography scripts. The re-markers marked segments of the scripts on screen, as in live marking. Analyses compared the marks awarded by the senior examiners during the re-marking exercise with prime marks from live marking.

Three indices of marking reliability are reported for the on-screen marking of essays. These are correlation coefficients, absolute mark differences and level of agreement. Values for all of the indices are comparable with those reported previously in the literature for the marking of essays, either on paper or on screen. After accounting for the effects of the leniency/severity profiles of individual re-markers and groups of prime markers, the difference between the two sets of marks for the essays was always less than 1 mark and, therefore, not considered to be educationally significant, although statistical significance was found. At the level of whole scripts the three indices of reliability all indicate high marking reliability. The correlation coefficient was 0.88; 78.6% of scripts showed mark/re-mark agreement within 5% of the maximum mark for the question paper (120 marks); and, the mean absolute mark difference as a percentage of the maximum mark was 3.7%.

The report concludes that further investigations into the on-screen marking of essay-type responses are needed. However, experimental studies could be augmented with operational data. These data are rich in information concerning the double marking of essay-type responses and the relationships between the timing of marking and error rates. Such information could be used as the basis of advice to examiners on best practice when marking.

KEYWORDS

on-screen marking, marking reliability, essay-type responses, essays

1. BACKGROUND

AQA is in the process of migrating the marking of scripts from conventional paper-based marking to on-screen marking or e-marking. This process began in 2004 and the intention is that it will be complete by the summer of 2012 (AQA, 2009). To meet this objective of the Strategic Action Plan, assessments that use extended or essay-type responses (essays) must be brought into the e-marking fold. Where essays are necessary for the validity of the assessment, that validity should not be compromised through migration to e-marking.

E-marking takes place within the CMI+ e-Marker® application. Examiners use the application to mark on screen the scanned images of responses to items from within scripts. The unit of on-

screen marking is an item, not a whole script. In the last year and based on the results of recent studies (Fowles, 2006a, 2008), CMI+ e-Marker® has been adapted to cope with the marking of essay-type responses. The purpose of these adaptations was to transfer those aspects of paper-based marking that are important to the marking of essays. The adaptations include the functionality to add annotations and comments to responses. The comments may be from a set that is pre-loaded before standardisation and from which an examiner selects appropriate comments by dragging and dropping. Alternatively, an examiner is able to key in comments that are specific to a response. Another adaptation is the inclusion of thumbnails from which an examiner is able to tell at a glance how many pages an essay runs to. The thumbnails act as a visual cue that places a psychological limit on the amount of scrolling required. This is the equivalent of an examiner flicking forwards and backwards through the pages of a paper script to familiarise themselves with a response.

A pilot phase is underway in which a small number of question papers containing essays are marked on screen using the freshly adapted CMI+ e-Marker® application. This piloting provides an opportunity to investigate the reliability of on-screen marking of essays and the question papers that contain them. The current report presents the results of a recent opportunistic marking reliability study of the on-screen marking of an AS level Geography question paper that contains essays. The study also included a survey of examiners' opinions about their experiences of on-screen marking of essays which will be reported formally at a later date.

1.1 Marking reliability studies of on-screen marking of essays

The transfer of marking of essays from paper to on-screen is a fairly recent phenomenon. Initial studies compared marking in the paper-based mode with marking on screen either for single essays (see, for example, Johnson, Nádas, Bell, & Green, 2009) or complete scripts (*inter alia* Fowles, 2006a). Johnson *et al.* (2009) reported no change in the leniency/severity of marking of examiners who marked, on paper and on screen, 180 responses to one essay (maximum mark 30) from a GCSE examination in English Literature. Their conclusion was qualified by the authors noting that there was greater variation in the examiners' on-screen marking compared with their marking on paper. This was attributed to between-marker variation and therefore most likely due to the subjective nature of marking essays. Fowles (2006a) compared the marking reliabilities of 35 paper scripts and 40 electronic scripts from a GCSE English specification A Higher tier examination. Correlation coefficients at script level for the on-screen marking, comparing the 'true' marks of the Principal Examiner with those of the five examiners, ranged from 0.591 to 0.798. Fowles observed an 'exaggeration' of examiners' marking tendencies towards leniency or severity in the on-screen mode when compared with their marking on paper. This is counter-balanced by the segmented nature of marking on screen by which the leniency of one examiner may be cancelled out by the severity of another examiner within a script. There was also evidence that the examiners taking part in the study experienced greater difficulty in transferring their marking skills to the essays in section B (maximum mark 27) of the scripts than to the short structured responses (maximum marks ranging from 3 to 8) in section A. Some of this difficulty was ascribed to the limitations placed on examiners' ability to manipulate the on-screen responses by the CMI+ e-Marker® application. Attention was drawn to scrolling, annotating and inserting comments. Overall, and given that improvements have been made to the e-Marker® application, the results from such studies indicated a successful transfer between modes of marking for essays.

Fears that some of the features of e-marking would prevent the successful migration of the marking of essays on screen have not been realised. The next question to receive attention is: do these same features of e-marking reduce the mark/re-mark reliability for marking essays on

screen? The literature contains very few reports of studies into the reliability of on-screen marking of essays *per se*. Cheung and Chan (2009) looked at the double marking of standardised scripts in on-screen marking of the written component of a low stakes test[1]. They reported correlation coefficients between the ratings from nine examiners and a consensual 'true' rating from a small committee of experts. The correlations of the ratings for content (rating scale 0 to 4) ranged from 0.869 to 0.957 and for the ratings of language (rating scale 0 to 3) the correlations lay in the range from 0.858 to 0.934.

Fowles (2009) reported the results of the double marking on screen of items from a sample of 17 scripts from a GCSE History examination. The question paper provided candidates with the choice of one from three topics in each of two sections. Each topic contains a number of part questions with maximum marks ranging from 5 to 15. Candidates are required to respond to one essay question worth 15 marks. In considering the reliability of the on-screen marking of this paper, the marks of 'amalgamated' examiners were compared with the Principal Examiner's 'true' mark. 'Amalgamated' described virtual examiners whose quota of experimental marking was constructed by systematically adding item level marks from a pool of such marks until the requisite number of script totals for each virtual examiner was reached. This was necessary as the examiners participating in the study did not mark all of the items from each of the original sample of 24 scripts. Correlation coefficients at script level between the marks of the amalgamated examiners and the 'true' marks ranged from 0.838 to 0.926. The mean absolute mark difference (AMD) across all amalgamated examiners and available scripts was 6.6% of the maximum mark (60 marks) for the question paper. Unfortunately, the conclusions from this study were equivocal due to attrition of both participants and sample.

1.2 A 'true' mark for segmented marking

Two definitions of 'true' mark are most frequently used in the reporting of reliability indices (Meadows & Billington, 2005). The *hierarchical mark* is given by the Principal Examiner who sets the question paper and mark scheme and is responsible for training examiners in how to mark the question paper. This mark is subject to any error present in the Principal Examiner's marking (Cresswell, 1982). The second definition of 'true' mark is the *consensual mark*. Such a mark may be arrived at by taking the mean of the marks awarded by all markers of a script. This is similar to the 'true' mark of classical test theory which is the result of the pooled judgement of an infinite number of markers (Spearman, 1927 as cited in Meadows and Billington (2005)). Alternatively, the consensual mark may be arrived at through discussion within a team of markers, as in the study of Cheung and Chang (2009).

Marking reliability studies using these definitions of 'true' mark usually considered the marking of whole scripts by individual examiners (see the review of marking reliability by Meadows and Billington (2005)). As Pinot de Moira (2009) and Fowles (2009) point out, the concept of a hierarchical 'true' mark emanating from the Principal Examiner may no longer be suitable for the segmented marking that takes place on screen. This study proposes a modified definition of the 'true' mark that is more appropriate to on-screen, segmented marking. If the Principal Examiner is the source of a *sole hierarchical mark*, then a few senior examiners marking item by item may provide a *group hierarchical 'true' mark*. This modified definition retains the element of seniority within the examining community alongside a representation of the practicality of marking that is closer to the reality of e-marking. Use of this group hierarchical mark as the basis of reporting

[1] The written component is part of the Primary 6 English Language in the Territory-wide System Assessment in the Hong Kong Special Administrative Region. This is part of a survey by assessment of standards in schools. The results of the survey inform the management of resources in the educational system.

reliability indices assumes that the members of the group are all marking to the same standard. Unlike Spearman's definition, the group hierarchical mark is not a mean mark. It represents a way of marking in which a team of examiners contributes marks to a script.

1.3 The research question

The current investigation used the definition of a group hierarchical 'true' mark to measure the mark/re-mark reliability of on-screen marking of essay-type responses and the question paper that contains them. The investigation addressed the research question:

Is the mark/re-mark reliability of the on-screen marking of essay-type responses and the question paper containing them at least as reliable as the mark/re-mark reliability of similar responses and question papers, whether on-screen or paper-based?

Responses to this research question were formulated through comparisons of the **prime marks** awarded during live marking with the **re-marks** provided by a small group of senior examiners during a re-marking exercise. Marking was segmented and on-screen in both live marking and the re-marking exercise; none of the examiners marked whole scripts at any time.

2. METHOD

The modified version of CMI+ e-Marker® was piloted with four A Level units in the January 2010 examination series. GEOG1, containing eight essay questions to the three essay questions of the other three units, was an opportunistic selection based on operational decisions and the relative experience of e-marking amongst the examining teams. The 'many scripts, few examiners' approach was used to investigate the reliability of the on-screen marking of GEOG1, in particular those items requiring an essay-type response. This approach, used in earlier marking reliability studies (Cresswell, 1982; Fowles, 2006b; Murphy, 1978a; 1978b; Newton, 1996) is suited to studies of short duration that focus on mark/re-mark reliability. The contrasting approach, 'few scripts, many examiners', used by, for example, Delap (1993) and Fearnley (2006), lends itself to a diagnosis of problems within a question paper and mark scheme.

2.1 The question paper

GEOG1 is an AS unit in the new specification for A Level Geography. It consists of two sections each with a maximum mark of 60, giving the entire question paper a maximum mark of 120. Section A is concerned with physical geography and Section B with human geography. There are four questions in both sections. The first question in a section is compulsory and candidates then choose one question from the remaining three optional questions. Each of the eight questions on the question paper has a maximum mark of 30 and consists of a number of part-questions requiring short structured responses worth between 4 to 7 marks and one part-question requiring an essay worth 15 marks. All items are expert marked.

2.2 Selection of examiners and scripts

The structure of the examining team for GEOG1 is slightly unusual. There is one team of examiners for each section of the question paper. (It is thought that individual examiners are unlikely to have the volume of subject knowledge needed to mark successfully in both physical geography and human geography.) At the top of the examining hierarchy is the Principal Examiner who oversees both teams of examiners. Each team has its own hierarchy, headed up by an Assistant Principal Examiner who manages a number of team leaders (three in the January 2010 examination series) who each lead a group of between three and five examiners. All members of the examining team were rated as grade A in the previous summer examination

series. All of the team leaders and junior examiners are expected to be team leaders in the next summer examination series.

The Principal Examiner was unavailable to participate in the re-marking exercise, removing the possibility of setting a sole hierarchical 'true' mark. Three of the four senior examiners in each team took part in the re-marking of a sample of 173 scripts. Thus, the reliability considered was group hierarchical, with multiple senior examiners who marked items on screen in the segmented manner. Emphasis was placed on selecting senior examiners only for participation in the re-marking exercise rather than a random group of examiners regardless of status within the examining hierarchy.

Initially, a random sample of 180 whole scripts was taken from a pool of scripts selected from the population of whole scripts that were marked on screen. There was some attrition of the sample due to the removal of seven scripts that were later identified as standardisation scripts. Selection of scripts into the pool in preparation for sampling was based on three criteria. Firstly, candidates had responded to all 17 items that made up a complete script, regardless of the route they pursued through the optional questions in the question paper. The second criterion was that only scripts containing responses to either the Cold Environments question (q2) or the Coastal Environments question (q3) from Section A and either the Energy Issues question (q7) or the Health Issues question (q8) from Section B could be in the sample. The questions about Hot Desert Environments and Their Margins (q4) and Food Supply Issues (q6) proved to be unpopular with candidates. Only one senior examiner was standardised to mark q4 and so marked all 118 responses. Seven senior and junior examiners marked the 620 responses to q6.

A consequence of applying these first two criteria was that the range of marks for whole scripts was reduced from 0 to 103 for the entire e-marked cohort to 30 to 78 for the sample. This reduced mark range still covered all five grades due to the low mark for the grade E boundary (33 out of 120) and severe mark compression. Unsurprisingly, the grade distribution for the sample shows a higher performing group of candidates at all grades in comparison with the entire cohort; see Table A1.

The third criterion was that the re-markers had not marked any of the items with maximum marks of 6 and above during the live marking period. This included the six essays (q1d, q2c, q3c, q5d, q7c and q8c), items with a maximum mark of 7 (q2bii, q3b and 5c) and items with a maximum mark of 6 (7aii and 8aii). With a relatively short time between the live marking and the experimental re-marking it was possible that the re-markers might have residual memories of some of the extended responses they had prime marked. Such memories might have affected the re-marking of items. This memory-effect was not thought to apply to the lower tariff items as the speed with which such items are marked is higher, leaving less time for individual responses to be retained in memory.

The sample of scripts was checked for being representative of the cohort in terms of centre type, gender and optional route through the question paper. The results of comparisons between the characteristics of the entire cohort and the sample are shown in Appendix A. The proportions of scripts with optional questions 2 & 7 and 2 & 8 are the only characteristics in which the sample differs greatly from the entire cohort; see Table A4. After removing those scripts with responses to optional questions 4 and 6, 14.2% of the entire cohort chose questions 2 & 7 and 20.5% chose questions 2 & 8. The equivalent percentages in the sample were 21.4%

and 16.2% responding to questions 2 & 7 and questions 2 & 8, respectively. Otherwise, the sample is considered to be representative.

2.3 Re-marking of sampled scripts

The re-marking of the sample of scripts took place under conditions similar to those that pertain during live marking. The six participating senior examiners re-marked items using the CMI+ e-Marker® application in the same environments in which they undertook their live marking. The majority of the essay responses in the sample were two to three pages in length. Examiners needed to scroll through these responses to mark them. The re-marking exercise was completed over five days and within three weeks of the end of the live marking period. This should have meant that the mark scheme was still fresh in the minds of the re-markers, as was familiarity with the electronic marking application. In addition, the items presented to the re-markers were clean, so there were no marks or comments from the prime markers which could have influenced the re-markers (Murphy, 1979). Presenting re-markers with clean items was simple to do within CMI+ as the marks and comments from the prime marker were stored in a separate transparent layer that sat on top of the image of the item.

Instructions to the re-markers included two requests. Firstly, to mark across the entire range of items in the sections of the question paper for which they were standardised. As with live marking, this was difficult to control remotely when the re-markers were marking at different times and rates. Consequently not all of the re-markers on a section of the question paper contributed to the marking of every item. The numbers of items that each re-marker marked are shown in Tables A5 and A6. The second request was to use the amount of annotation that enabled them to comprehend a response and award what they felt to be the appropriate mark for each response. They were not required to annotate for *post hoc* judgemental processes as there was no intention to alter candidates' marks based on the re-marking exercise. Reduced annotation decreased the ecological validity of the study somewhat.

2.4 Analyses

The following indices of marking reliability were used when analysing the data.

1. The correlation between prime marks awarded during live marking and re-marks awarded by the small group of senior examiners during the re-marking exercise. This is a measure of the level of agreement over the ranking of scripts and responses to items.
2. The degree of agreement between prime marks from live marking and re-marks from the re-marking exercise, either complete or within 10% of the maximum mark.
3. The absolute mark difference (AMD) between the prime marks from live marking and re-marks from the re-marking exercise. This is a measure of marking accuracy or how divergent prime marks are from the re-marks.
4. The relative severity or leniency of the prime marks in comparison with the re-marks.

3. RESULTS

It is inappropriate to apply correlational analyses to variables with restricted ranges. Doing so may reduce the statistical power of the analyses leading either to effects being missed (Type II errors) or to the detection of spurious effects (Type I errors). Therefore, the current study limits the reporting of correlations to scripts, sections of the question paper and the essays (maximum marks of 120, 60 and 15, respectively).

The results section is in two parts. The first part presents the results of the investigation into the marking reliability of the essay-type responses. Part two reports the reliability indices at script

and section levels. The values for descriptive statistics and correlation coefficients, AMDs (means, s.d.s and as a percentage of the maximum mark) and the percentage agreement between prime marks and re-marks for all items, both sections and scripts are in Appendix B. Values from these tables are used in the main text to illustrate points of interest. On occasion analyses of variance (ANOVAs) were used to probe the data. Unless stated otherwise assumptions of normality, homogeneity of variances and sphericity were met. The ANOVA summary tables are in Appendix C.

3.1 Reliability of marking of the essay-type response items

3.1.1 Correlation coefficients: rank ordering of responses

Values of the indices of marking reliability for the essays are shown in Table 1. Half of the essays showed strong positive relationships between the prime marks from live marking and the re-marks from re-marking by senior examiners only (column 3 of Table 1 for q3c, q5d and q8c). The three remaining essays showed moderate positive relationships between the two sets of marks. The values of the correlation coefficients are encouraging in light of the low mean marks (columns 6 and 8 in Table 1), which are, on average, 38.9% and 35.8%, of the maximum mark for prime marks and re-marks, respectively.

Table 1: Correlation coefficients and absolute mark differences between prime marks and re-marks and descriptive statistics for the essays (maximum mark 15)

		Correlation	AMD as % of max. mark		Prime mark		Re-mark	
Item	**N**	**coefficient[2]**	**Mean**	**s.d.**	**Mean**	**s.d.**	**Mean**	**s.d.**
1d	173	0.593	13.2	10.4	6.57	2.61	5.68	2.62
2c	65	0.623	11.2	8.1	6.09	2.02	5.86	2.59
3c	108	0.854	9.5	9.8	4.97	3.46	4.01	3.12
5d	173	0.702	11.9	9.7	6.16	2.93	5.69	2.93
7c	85	0.691	10.4	9.2	5.41	2.39	4.46	2.35
8c	88	0.748	10.9	7.7	5.78	2.82	6.49	2.31

3.1.2 Absolute mark differences: accuracy/divergence of marking

The absolute mark difference is a measure of the divergence of the prime mark awarded during live marking from the re-marks given by the small group of senior examiners during the re-marking exercise. The compulsory essays in each section showed the greatest divergence between the two sets of marks. Q1d and q5d had mean AMDs of 13.2% and 11.9% of the maximum mark for essays, respectively (column 4 of Table 1). In their respective sections of the question paper these essays also had the largest spreads of AMDs (column 5 of Table 1). Such differences were attributed to differences between the questions and their mark schemes (the question paper and mark scheme are in Appendix D for reference).

3.1.3 Severity/leniency of marking

The third index of reliability was the comparative severity or leniency of marking. The prime marks from live marking were lenient in comparison with the senior examiners' re-marks for five of the six essays, the exception being q8c (columns 6 and 8 in Table 1). As might be expected when marking is subjective, two-tailed paired sample t-tests showed that the difference between

[2] For all correlations $p < 0.001$.

the means for prime marks and re-marks was statistically significant for five of the six essays[3]. However, these differences were less than 1 mark for each of the essays.

The reliability of the on-screen marking of the essays appears to be reasonably high considering the values of the three indices outlined above. These indices are insufficient evidence of high marking reliability because the mean marks from the two marking occasions are statistically significantly different. It would be informative to remove the effects of the leniency/severity profiles of individual markers or groups of markers. To this end, 3-way mixed analyses of variance were carried out on the data from the essays. Mixed ANOVAs were applied to each section of the question paper rather than the entire question paper because of the wholly different examining teams. The within-subjects variable was the mark awarded to the essays on the two marking occasions: live marking by junior and senior examiners and re-marking by seniors only. The ANOVA controlled for two between-subjects factors. The first of these factors was the individual re-marker. There were three re-markers for each section, each of whom had their own tendency to leniency/severity in marking when compared with other examiners. The second factor was the status of the prime marker. The sample of scripts was not selected to be representative of the live marking of all examiners on each item. However, it was possible to identify prime markers by their status within the examining hierarchy during live marking. In this case, identification was limited to whether a prime marker was a junior or senior examiner. The latter description applied to team leaders and above.

The difference between the prime marks from live marking and the re-marks for the essays in section A was unlikely to have arisen by chance in this sample ($F(1,340) = 7.90$, $p = 0.005$, $eta^2 = 0.023$, power = 0.800). Although the difference between the prime marks (M = 5.98, SD = 2.89)[4] and the re-marks (M = 5.19, SD = 2.89) is statistically significant after controlling for the marking tendencies of individuals and groups, the effect size is small and the difference between means is less than 1 mark. The situation for the essays in section B was similar but not statistically significant ($F(1,340) = 1.332$, $p = 0.249$, $eta^2 = 0.004$, power = 0.210)[5]. The difference between the mean mark for prime marks from live marking (M = 5.88, SD = 2.79) and the mean mark from the re-marking exercise (M = 5.59, SD = 2.74) was small and likely to be due to chance for this sample.

The between-subjects factor of status of prime marker did not exert a statistically significant effect on the difference between prime marks and re-marks for essays from either section A or section B ($F(1, 340) = 0.58$, $p = 0.446$, $eta^2 = 0.002$, power = 0.118 and $F(1, 340) = 0.71$, $p = 0.399$, $eta^2 = 0.002$, power = 0.134, respectively). Despite this, the plot of estimated marginal means, shown in Figure 1, illuminates a trend in the marking of GEOG1. The filled symbols are the mean marks for the essays from section A and the open symbols are the mean marks for the essays from section B. Within the sample, the mean prime mark from junior examiners during live marking on section A was 6.10. On re-marking by the senior examiners this decreased by 0.64 marks to 5.46. Again, within the sample, the mean prime mark awarded by senior examiners during live marking on section A was 5.48. During the re-marking exercise the

[3] q1d: $t(172) = -4.925$ $p < 0.001$; q2c: $t(64) = -0.900$ $p = 0.372$; q3c: $t(107) = -5.534$ $p < 0.001$; q5d: $t(172) = -2.721$ $p = 0.007$; q7c: $t(84) = -4.714$ $p < 0.001$; q8c: $t(87) = -3.511$ $p = 0.001$

[4] Unadjusted mean marks are quoted here, but they are similar to the adjusted values of the estimated marginal means.

[5] The Greenhouse-Geisser correction was applied for both sections A and B as the assumption of sphericity was not met. For the essays in section B the assumption of homogeneity of variances was not met for the prime marks ($F (df1 = 5, df2 = 340) = 2.802$, $p = 0.017$).

small group of senior examiners decreased this by 0.37 marks to 5.11. The means for prime marks and re-marks for the essays in Section B showed a similar pattern.

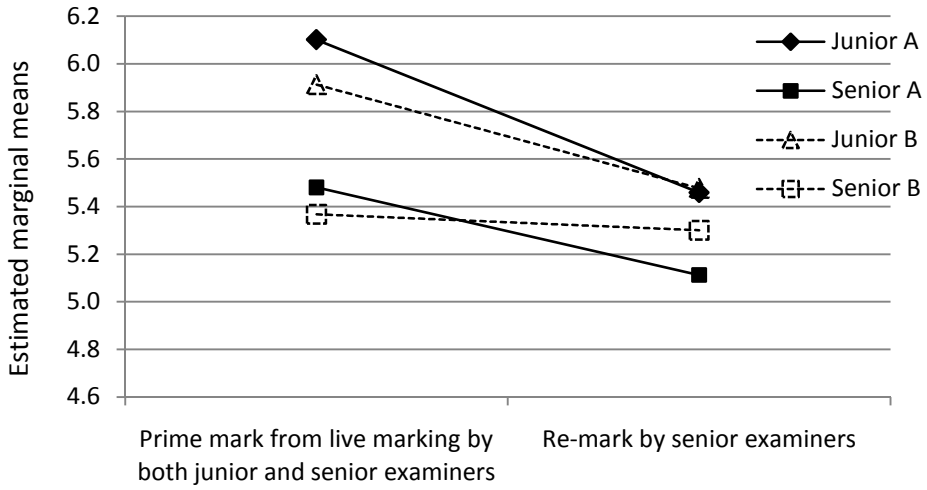


Figure 1: Estimated marginal means for prime marks from live marking and re-marks awarded by senior examiners on essays in sections A and B of GEOG1. This plot shows essays that were prime marked by junior and senior examiners during live marking.

3.2 Marking reliability at script and section levels

3.2.1 Correlation coefficients: rank ordering of scripts

The correlation coefficients between the prime marks and re-marks at both whole script and question paper section levels were equal to or greater than 0.80 (column 3 in Table 2). They were all statistically significant ($p < 0.001$). This index of reliability indicates that both sets of markers were able to rank order to similar extents whole scripts and the separate sections within the question paper.

3.2.2 Absolute mark differences: accuracy/divergence of marking

The absolute mark differences (column 4 in Table 2) showed that the accuracy of marking over whole scripts was reasonable. A mean AMD of 4.45 was 3.7% of the maximum mark for the question paper. The mean AMD for each of the sections were approximately 3 marks, or 5% of the maximum mark for the section (60 marks).

Table 2: Correlation coefficients and AMDs between prime and re-marks and descriptive statistics at whole script and section levels

	Max.	Correlation	AMD as % of max. mark		Prime mark		Re-mark	
Item	**mark**	**coefficient**	**Mean**	**s.d.**	**Mean**	**s.d.**	**Mean**	**s.d.**
Section A	60	0.85	5.2	4.0	25.09	6.67	23.95	6.93
Section B	60	0.80	5.3	4.2	25.83	6.02	25.53	6.63
Whole script	120	0.88	3.7	2.8	50.92	10.60	49.48	11.42

Figure 2 shows the cumulative percentage of whole scripts that were awarded a prime mark during live marking that was less than or equal to a specific absolute mark difference away from

the re-mark (as a percentage of the maximum mark) awarded by the small group of senior examiners. It also shows the same information for each of the sections on the question paper. At script level there was complete agreement between the prime marks and the re-marks for 5.2% of the sample. There is agreement within 10% of the maximum mark for the whole script (120 marks) for almost all of the scripts (96.5%). Due to the low grade E boundary (33 out of 120) and severe mark compression that produced ranges of 7 marks for grades B, C and D and a range of 6 marks for grade E, it was deemed more appropriate to consider the agreement within 3.5 marks or half a grade width (2.9% of the maximum mark for the script). In this case there is agreement for 53.5% of scripts. This means that, following the re-marking exercise, just under half of the scripts received a mark that had the potential to change the grade of the script. The actual proportion of scripts with a grade change due to using re-marks was 51.5% (see Table 3). 8.1% of scripts changed by more than one grade on re-marking.

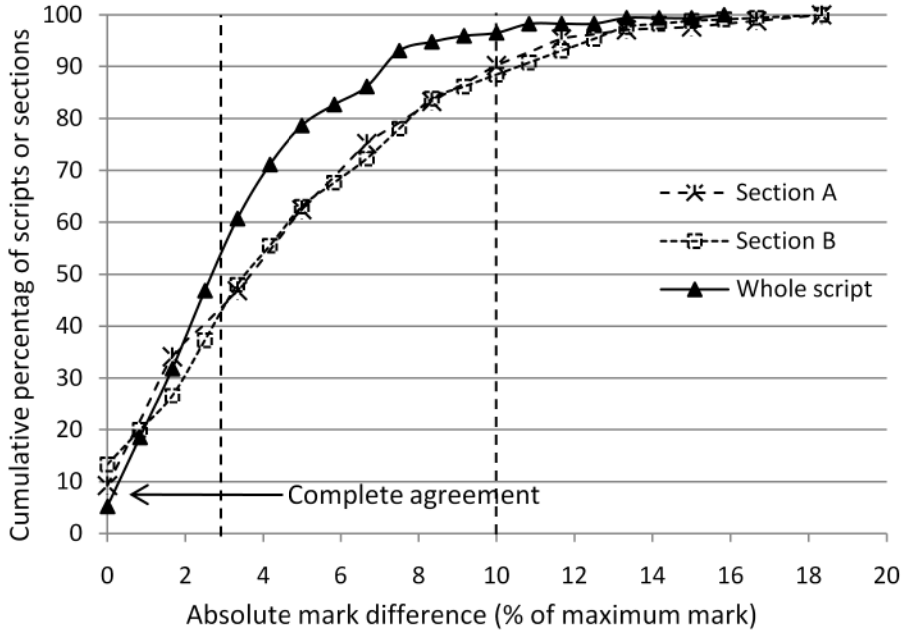


Figure 2: The percentage cumulative frequency of scripts and sections awarded prime marks that are less than or equal to a specific absolute mark difference (as a percentage of maximum mark) away from the re-mark

Table 3: Number of potential grade changes due to the re-marking exercise

	Grade changes from prime marking (% of scripts)						
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**
All re-marks	0.58	4.62	28.32	48.55	15.03	2.89	0.00

Sections A and B showed similar patterns of agreement with 90.2% and 88.4% of prime marks for sections A and B, respectively, agreeing with the re-marks within 10% of the maximum mark for a section (60 marks). For absolute mark differences greater than 2% of the maximum mark, the agreement between prime marks and re-marks was lower for the marks awarded to the sections within the question paper than the marks awarded for whole scripts. That the degree of agreement between prime marks and re-marks decreased when moving from whole script to sections indicates the actual differences between prime marks and re-marks tended to cancel

each other out on aggregation to the whole script level. Adding up the mean AMDs for sections A and B in Table 2 confirms this. The addition gives a total of 6.29 marks, approximately one and a half times greater than the mean AMD for whole scripts.

3.2.3 Severity/leniency of marking

A comparison of the means for prime marks from live marking and for the re-marks from senior examiners indicated that the prime marking was lenient. Evidence from paired sample t-tests showed this leniency in marking was statistically significant for whole scripts ($t(172) = 3.506$, $p = 0.001$, two-tailed) and section A ($t(172) = 3.999$, $p < 0.001$, two-tailed), but not for section B ($t(172) = 0.970$, $p = 0.333$, two-tailed).

Two factors were identified that may have caused the differences between the mean marks at script level. The first factor was the route taken through the question paper by a candidate. In this study 'route' is taken as a measure of how difficult the examiners found the scripts to mark during live marking and is therefore ascribed the role of independent variable. The difficulty in marking could be due to the questions themselves being difficult, the mark schemes being deficient or the erratic nature of the responses leading to inconsistent marking. The second factor was the number of markers contributing to the prime mark during live marking. Between 8 and 16 examiners contributed to the prime marking of the scripts in the sample, with the average number of contributing examiners being 13. In the re-marking exercise all six of the participants contributed to the marking of the majority of scripts (71.1%). Of the remaining scripts, 26.0% (45 out of 173) were re-marked by five participants and 2.9% (5 out of 173) were re-marked by four participants.

A 3-way mixed analysis of variance using script mark as the within-subjects factor was carried out on the data. The ANOVA controlled for the number of examiners contributing to the prime marking of each individual script and the route through the question paper. The numbers of examiners contributing to the prime mark were split into five groups of approximately equal size. The difference between the means of the prime marks ($M = 50.92$, $SD = 10.60$) and re-marks ($M = 49.48$, $SD = 1.42$) at script level was statistically significant ($F(1,153) = 7.88$, $p = 0.006$, $eta^2 = 0.049$, power = 0.797). However, the effect size was small, as was the difference between the means of the two sets of marks at 1.44 marks (or 1.2% of the maximum mark for a script)[6]. Route exerted a statistically significant effect on the difference between prime marks and re-marks ($F(3,153) = 2.76$, $p = 0.044$, $eta^2 = 0.051$, power = 0.659). The number of examiners contributing to the prime mark did not have a significant effect on the difference between the prime marks and the re-marks ($F(4,153) = 0.57$, $p = 0.685$, $eta^2 = 0.015$, power = 0.186).

4. DISCUSSION

4.1 Reliability of on-screen marking of essays

The reliability indices for the mark/re-mark reliability of the essay-type responses in GEOG1 indicated reasonably high on-screen marking reliability. Prime markers awarding marks during live marking showed moderate to high levels of correlation with the small group of senior examiners who took part in the re-marking exercise. The correlation coefficients for the marking of the essays ranged from 0.59 to 0.85. These values compared favourably with correlation coefficients of 0.44 and 0.88 reported by Fowles (2006b) for the paper-based marking of essays (maximum mark 13 or 18, plus 9 marks for SPG) within scripts from both Foundation and Higher

[6] The estimated marginal means were 51.87 (SE = 1.00) for prime marks and 50.44 (SE = 1.07) for re-marks. The Greenhouse-Geisser correction was applied as the assumption of sphericity was not met.

tiers of GCSE English specifications A and B. The absolute mark differences as a percentage of the maximum mark (15 marks) ranged from 9.5% for q3c to 13.2% for q1d. Again, these compared favourably with previous research showing a range of differences from a sole hierarchical true mark of between ±11 for a single essay with a maximum mark of 30 taken from a GCSE English Literature question paper that was marked on screen (Johnson *et al.*, 2009). Fowles (2006b) also reported mean absolute mark differences for the paper-based marking of essays within scripts from both Foundation and Higher tiers of GCSE English specifications A and B. These ranged from 3.9% to 14.6% of the maximum mark for the essay (excluding SPG) and showed some dependence on the number of marks in a marking level.

In the current investigation analyses of variance of the marking of the essays in each section of the GEOG1 question paper showed statistically significant differences between the prime marks from live marking and the senior examiners' re-marks for section A, but not for section B. This was after controlling for the presence of more and less accurate markers amongst both the re-markers and the two groups of junior and senior prime markers. However, the differences in mean marks between the two marking conditions did not exceed 1 mark and may, therefore, be considered not educationally significant.

During live marking of the essays, the junior examiners marked more leniently than did the senior examiners on both sections of the question paper. The senior examiners who participated in the re-marking exercise reduced the prime marks of the junior examiners from live marking by approximately twice as much as they reduced the prime marks of the senior examiners. As the small group of senior examiners re-marked clean responses this differential in severity was not influenced by the re-markers knowing who had previously marked a particular response. This evidence points to the on-screen marking of essays being reliable despite the variation in the leniency/severity of marking of individual markers and groups of markers.

The observation that the junior examiners tended to mark the essays leniently in comparison with the senior examiners is unusual in itself. The literature indicates senior markers are usually more willing to give the benefit of the doubt to responses and so can appear lenient in comparison with less experienced junior examiners (Ruth & Murphy, 1988; Weigle, 1999). Experience in these studies was measured in terms of number of years of teaching. Royal-Dawson and Baird (2006) used the same measure of experience in their investigation of the effects of markers' backgrounds on the marking of Key Stage 3 English. They found that on a sub-test for reading the group of examiners with the most teaching experience tended towards leniency. In the current investigation, information about the length of teaching experience of the examiners was not available. The mean length of examining experience was 5 years and 11 years for junior and senior examiners, respectively. However, Royal-Dawson and Baird (2006) have pointed out that this measure of experience is confounded because reliable examiners are employed year after year and poor markers are not. Within AQA's operational parameters, all of the examiners who marked GEOG1 in January 2010 were considered to be reliable. They were rated grade A in the previous examination series and are expected to be team leaders in the next series. The difference between the prime marks awarded by junior and senior examiners during live marking was systematic. This suggests that the marking standard was not disseminated fully by the seniors to the juniors during standardisation.

There is an additional confound because this new A Level Geography specification is the result of moving from six units to four units and from two specifications to one specification. The examiners have different backgrounds in terms of the legacy units and specifications that they

marked prior to joining the GEOG1 examining team in the June 2009 examination series. The differences in marking background may inform the 'ideological biases' of individual examiners (Husbands, 1976). As the January 2010 examination series was only the second in which GEOG1 was examined, such biases may be smoothed out over time provided the marking standard is communicated to all examiners.

4.2 Reliability of on-screen marking of whole scripts

The correlation coefficients for the on-screen marking/re-marking of GEOG1 were comparable with those measured for the paper-based marking of GCSE English components containing questions that required structured responses of varying lengths (Fowles, 2006b; Newton, 1996). Both Fowles (2006b) and Newton (1996) used a sole hierarchical true mark. The former reported coefficients of between 0.40 and 0.95, depending on the type of mark scheme used. The latter reported coefficients of 0.81 to 0.93. The coefficients at whole script level were also comparable with correlation coefficients of between 0.84 and 0.93 reported by Fowles (2009) for the segmented on-screen marking of GCSE History scripts, in which the marks of amalgamated examiners were compared with a sole hierarchical mark.

As a percentage of the maximum mark, the mean absolute mark differences compared favourably with previous studies. The values were 5.2%, 5.3% and 3.7% for section A, section B and whole script, respectively. Fowles (2009) reported mean absolute mark differences of between 5.4% and 7.7% of the maximum mark for a GCSE History question paper that was marked on screen.

A comparison of the curve for whole scripts with the simulated curves showing the probability that a total mark will be less than or equal to a given absolute difference away from the true mark shown in Figure 1 of Pinot de Moira (2009) is reassuring. The curve for GEOG1 whole scripts in Figure 2 falls between the simulated curves for GCSE Religious Studies and GCSE English A in Pinot de Moira's report. The two components considered by Pinot de Moira (2009) cover the extremes of the range of item types. The component from GCSE Religious Studies was comprised of short structured response items worth up to 8 marks and the component from GCSE English consisted of two essay response items, each worth a maximum of 27 marks. The component under study in this investigation, GEOG1, contains a mixture of short structured responses and essays. Therefore, it is logical that, in terms of agreement between prime marks and re-marks, it would sit between the two components investigated by Pinot de Moira (2009). This statement needs to be qualified by acknowledging that the data Pinot de Moira used were based on a sole hierarchical true mark, with the marks of examiners being compared with those of a Principal Examiner.

Delap (1993) reported the results of a diagnostic mark/re-mark reliability study in the paper-based mode on Papers 1 and 2 of a GCSE Geography specification. Essays of unknown maximum marks were included in these question papers. The question papers themselves had maximum marks of 53 for Paper 1 and 130 for Paper 2. On considering the individual items within these question papers with the examiners, Delap found that some of the levels of response mark schemes were being converted informally to points-based marked schemes and *vice versa* by individual examiners. A cursory overview of the mark scheme in the January 2010 offering of GEOG1 (Appendix D) reveals the potential for similar behaviour amongst examiners. For example, q3i lists eight valid points that are mark-worthy for an item that has a maximum mark of 4. Where there are more points than marks there is a tendency for greater variation in the marking to be observed (Bramley, 2009). In GEOG1 there is also a heavy reliance on subject knowledge in the mark schemes although the Assessment Objective of

knowledge and understanding accounts for 51% of the maximum mark. Examiners who lack familiarity with a type of environment or issue, may produce inconsistent marking (Delap, 1993). Thus, question difficulty may affect marking reliability. Here, 'question difficulty' includes problematic mark schemes, difficulty with the knowledge of subject content a question requires from an examiner, as well as the erratic nature of the responses.

4.3 Limitations of the study

There were a number of challenges to the integrity of this study. First and foremost is that the study focussed on only one question paper. This challenge is compounded by GEOG1, unusually, having two examining teams, one for each section of the question paper. This severely limits the generalisation of the findings to the marking of other components on screen. As Fowles (2009) pointed out, to be able to comment with authority on the reliability of on-screen marking of essay-type responses, the marking of more question papers should be studied.

The current investigation considered only quantitative indices of mark/re-mark reliability. No attempt was made to identify or quantify which features of essays are valid for examiners to mark on-screen. This is a topic that is starting to be addressed in e-marking (see, for example, Johnson & Greatorex, 2008) and is worthy of further study.

Having the re-markers use CMI+ e-Marker® during the study exactly as and where they used it during live marking meant that the re-marking was representative of on-screen segmented marking and lends ecological validity to the study (Chamberlain, 2008). Despite this there is no control for level of motivation which may be less than during live marking. The re-markers were aware that they were over-marking responses and that their re-marks would not result in a change to the final mark of any response.

This study used the CMI+ e-Marking® application with a database that was a copy of the one used in live marking, but that operated entirely separately from the live marking database. This *modus operandi* was advantageous. The marking reliability study was run close to the end of the live marking period to aid the retention of the mark scheme and familiarity with the e-marker application in the memories of the re-markers. Previously, such timing was considered unfeasible due to the potential for disruption of re-marking post results. The use of a separate database to store responses and marks also offers the ability to run concurrent studies that use one database per participant. Each database could be tailored either to the design of the study or to the measured marking tendency of the participant, or both. It is possible to direct responses to specific markers within the CMI+ e-Marker®, however the use of separate databases may offer a level of security beyond that of the direct responses functionality. The *caveat* here is that the quality of live marking should not be threatened, whatever experimental method is used.

5. CONCLUSIONS

This study reported the marking reliability indices of correlation coefficients, absolute mark differences and level of agreement from a live mark/re-mark exercise using the question paper from an AS Level unit, GEOG1. A small group of senior examiners re-marked on screen all of the items from 173 scripts to provide a group hierarchical true mark. This modified definition of the true mark provided a better representation of the segmented nature of on-screen marking. Using the group hierarchical true mark, reliability indices for the on-screen marking of essay responses were comparable with those reported for the marking of essay responses in other subjects, whether on paper or on screen. The marking reliability at whole script level was also

comparable with the marking reliabilities that were reported in previous studies of the marking of complete scripts.

There were statistically significant differences between the mean marks for the two marking occasions. These differences were considered to be not educationally significant (less than one mark for the essays and 1.44 marks for the scripts) after controlling for the marking tendencies of individual examiners and groups of examiners. It was also acknowledged that there is a complex mix of the difficulty of optional routes through the question paper, deficient mark schemes and the erratic nature of responses that may have some impact of the quality of marking.

There are two recommendations based on the contents of this report and the data produced as a by-product of the CMI+ system. Firstly, that consideration is given to the design and funding of a broader study of the reliability of the on-screen marking of essays. The current study was opportunistic in nature. Whilst it has demonstrated an acceptable reliability for the on-screen marking of essays, designed studies provide benefits in terms of controlling for a number of variables. Such studies may consider what features of essays examiners award marks to and whether these are valid or not. The second consideration is that a further project fully utilising operational data from the live marking of a number of question papers that contain essays should be undertaken. The operational data are rich in information, not only about double marking, but also about when marking is undertaken and error rates that might be associated with the timing of marking. This recommendation echoes those in Fowles (2009) and Hudson, Donahue, Rutt and Schagen (2007).

Claire Whitehouse
Senior Research Officer 05 May 2010
Z:\Research Dept\projects\long form marking\reports\emarking essay types v14.docx

REFERENCES

AQA. (2009). *Strategic Action Plan AQA Moving to 2013*. Retrieved 12 April 2010, from http://intranet/pdf/action-plan.pdf

Bramley, T. (2009) Mark scheme features associated with different levels of marker agreement. *Research Matters,* v*June* n8 pp. 16-23.

Chamberlain, S. (2008). *Do Marking Reliability Studies have Validity?* (No. RPA_08_SC_RP_052). Manchester: AQA.

Cheung, K. M. A., & Chang, R. (2009, September) *Investigating Reliability and Validity in Rating Scripts for Standardisation Purposes in Onscreen Marking.* Paper presented at the 35th IAEA Annual Conference, Brisbane, Australia.

Cresswell, M. J. (1982). Reliability of marking. A study of five June 1981 examinations and a re-analysis of some previous data (No. RPA_82_MC_RAC_221). Guildford: SEG.

Delap, M. (1993). Marking reliabiity study in GCSE Geography (1163). A Study Of The Summer 1992 GCSE Examination, Papers 1 and 2 (No. RPA_93_MD_RAC_610). Guildford: SEG.

Fearnley, A. J. (2006). An investigation of targeted double marking: a report organised by the AQA as part of an NAA series of investigations into the enhanced quality of marking (No. RPA_06_AJF_RP_008). Manchester: AQA.

Fowles, D. (2006a). How well does marking in GCSE English transfer to marking using CMI+ with annotation? (No. RPA_06_DEF_RP_047). Manchester: AQA.

Fowles, D. (2006b). *How reliable is marking in GCSE English?* (No. RPA_06_DEF_RP_048). Manchester: AQA.

Fowles, D. (2008). Using the think aloud method to investigate electronic marking of long form answers (No. RPA_08_DF_RP_020). Manchester: AQA.

Fowles, D. (2009). A concurrent approach to estimating the reliability of electronic marking of long form answers (No. RPA_09_DEF_RP_070). Manchester: AQA.

Husbands, C. T. (1976) Ideological bias in the marking of examinations. *Research in Education,* v*15* pp. 17-38.

Johnson, M., & Greatorex, J. (2008) Judging Text Presented on Screen: implications for validity. *E-Learning,* v*5* n1 pp. 40-50.

Johnson, M., Nádas, R., Bell, J. F., & Green, S. (2009, November) *Marking essays on screen: an investigation into the reliability of marking extended subjective texts.* Paper presented at the AEA-Europe Annual Conference, Malta.

Meadows, M., & Billington, L. (2005). *A review of the literature on marking reliability* (No. RPA_05_MM_RP_05). Manchester: AQA.

Murphy, R. J. L. (1978a) Reliability of Marking in Eight GCE Examinations. *British Journal of Educational Psychology,* v*48* pp. 196-200.

Murphy, R. J. L. (1978b). Report of an Investigation carried out by the Research Unit of the Associated Examining Board. An investigation into the marking variations of individual examiners. (No. RPA_78_RM_RAC_067). Guildford: AEB.

Murphy, R. J. L. (1979) Removing the marks from examination scripts before re-marking them: Does it make any difference? *British Journal of Educational Psychology,* v*49* n1 pp. 73-78.

Newton, P. E. (1996) The Reliability of Marking of General Certificate of Secondary Education Scripts: mathematics and English. *British Educational Research Journal,* v*22* n4 pp. 405-420.

Pinot de Moira, A. (2009). Marking reliability & mark tolerances. Deriving Business Rules for the CMI+ Marking of Long Form Answer Questions (No. RPA_09_APM_TR_042). Guildford: AQA.

Royal-Dawson, L., & Baird, J.-A. (2006). *Is teaching experience necessary for reliable marking?* (No. RPA_06_LRD_RP_030). Guildford: AQA.

Ruth, L., & Murphy, S. (1988) *Designing writing tasks for the assessment of writing.* Norwood, NJ: Ablex Publishing Corp.

Spearman, C. E. (1927) The abilities of man, their nature and measurement. New York: Macmillan.

Weigle, S. (1999) Investigating Rater/Prompt Interactions in Writing Assessment: Quantitative & Qualitative Approaches. *Assessing Writing,* v*6* n2 pp. 145-178.

APPENDIX A

Table A1: Comparison of cohort and sample by grade distribution

	A	B	C	D	E	Number
Published results	17.6	34.4	55.1	75.9	88.6	4,696
Sample	22.5	44.5	63.6	85.5	97.7	173

Table A2: Comparison of cohort and sample by centre type

	Entire cohort		Sample	
Centre type	**Percentage**	**Number**	**Percentage**	**Number**
Secondary: Comprehensive/Middle	52.35	2,408	52.02	90
Independent	17.85	821	18.50	32
Sixth Form college	9.61	442	9.83	17
Secondary: Selective	8.22	378	8.09	14
FE establishment	3.37	155	3.47	6
Tertiary college	3.24	149	2.89	5
Secondary: Modern	2.52	116	2.89	5
City Academies	1.28	59	1.16	2
Other UK centres	0.93	43	1.16	2
Overseas centres	0.63	29	0.00	0
	100.00	4,600	100.00	173

Table A3: Comparison of cohort and sample by gender

	Entire cohort		Sample	
	Percentage	**Number**	**Percentage**	**Number**
Male	58.85	2,707	58.96	102
Female	41.15	1,893	41.04	71
Total	100.00	4,600	100.00	173

Table A4: Comparison of cohort and sample by optional route through question paper

	Entire cohort		Sample	
	Percentage	**Number**	**Percentage**	**Number**
questions 3 & 8	36.38	1,135	34.68	60
questions 3 & 7	28.94	903	27.75	48
questions 2 & 8	20.45	638	16.18	28
questions 2 & 7	14.23	444	21.39	37
	100.00	3,120	100.0	173

Table A5: Numbers of items marked by each re-marker on section A. ReMA1 indicates re-marker 1 on section A.

Item i.d.	Maximum mark	ReMA1	ReMA2	ReMA3
1a	4	60	61	55
1bi	3	60	62	54
1bii	3	61	62	53
1c	5	60	60	56
1d	15	0	106	70
2a	4	60	0	6
2bi	4	60	0	6
2bii	7	62	0	4
2c	15	60	0	6
3ai	4	48	0	62
3aii	4	50	60	0
3b	7	50	60	0
3c	15	60	50	0
Total items		**691**	**521**	**372**

Table A6: Numbers of items marked by each re-marker on section B. ReMB4 indicates re-marker 4 on section B.

Item i.d.	Maximum mark	ReMB4	ReMB5	ReMB6
5a	4	61	73	42
5b	4	60	74	42
5c	7	63	71	42
5d	15	60	74	42
7ai	4	28	17	42
7aii	6	28	17	42
7b	5	27	17	43
7c	15	28	17	42
8ai	4	47	0	42
8aii	6	25	22	42
8b	5	47	0	42
8c	15	2	43	44
Total items		**476**	**425**	**507**

APPENDIX B

Table B1: Descriptive statistics for all items. The rows containing information about the essay items are shaded.

			Prime mark		Re-mark	
Section	**Item**	**Max. mark**	**Mean**	**s.d.**	**Mean**	**s.d.**
A	1a	4	1.66	1.24	1.66	1.30
	1bi	3	1.65	0.93	1.51	1.00
	1bii	3	1.23	0.88	1.18	0.86
	1c	5	2.54	1.06	2.98	1.21
	1d	15	6.57	2.61	5.68	2.62
	2a	4	1.38	0.91	1.34	0.85
	2bi	4	1.71	1.14	2.11	1.15
	2bii	7	2.35	1.50	1.94	1.37
	2c	15	6.09	2.02	5.86	2.59
	3ai	4	2.67	0.90	2.71	0.92
	3aii	4	1.08	0.96	1.09	0.85
	3b	7	2.66	1.08	2.94	1.45
	3c	15	4.97	3.46	4.01	3.12
	Section A	60	25.09	6.67	23.95	6.93
B	5a	4	1.21	0.98	1.25	1.07
	5b	4	2.20	0.96	2.23	0.94
	5c	7	3.85	1.18	3.94	1.25
	5d	15	6.16	2.93	5.69	2.93
	7ai	4	1.38	0.98	1.26	0.98
	7aii	6	2.29	1.22	2.33	1.37
	7b	5	3.15	1.09	3.20	0.96
	7c	15	5.41	2.39	4.46	2.35
	8ai	4	2.22	1.04	2.11	0.98
	8aii	6	2.49	1.33	2.47	1.53
	8b	5	2.10	1.31	2.49	1.35
	8c	15	5.78	2.82	6.49	2.31
	Section B	60	25.83	6.02	25.53	6.63
GEOG1		120	50.92	10.60	49.48	11.42

Table B2: Comparison of re-mark and prime marks for GEOG1 (January 2010) using the three indices correlation coefficient, level of agreement and absolute mark differences (AMDs). The rows containing information about the essay items are shaded.

					Complete agreement		Agreement within range			AMD mean		AMD s.d.	
Section	**Item**	**Max. mark**	**N**	**Correlation coefficient**	**No.**	**%**	**Range**	**No.**	**%**		**as % of max. mark**		**as % of max. mark**
A	1a	4	173		90	52.0	1	159	91.9	0.56	14.0	0.64	16.0
	1bi	3	173		97	56.1	1	163	94.2	0.52	17.3	0.66	22.0
	1bii	3	173		124	71.7	1	167	96.5	0.32	10.8	0.56	18.7
	1c	5	173		74	42.8	1	137	79.2	0.79	15.8	0.79	15.9
	1d	15	173	0.593	22	12.7	2	118	68.2	1.98	13.2	1.55	10.4
	2a	4	65		51	78.5	1	64	98.5	0.23	5.8	0.46	11.5
	2bi	4	65		27	41.5	1	58	89.2	0.71	17.7	0.70	17.5
	2bii	7	65		27	41.5	1	53	81.5	0.78	11.2	0.80	11.4
	2c	15	65	0.623	11	16.9	2	52	80.0	1.68	11.2	1.21	8.1
	3ai	4	108		53	49.1	1	106	98.1	0.53	13.2	0.54	13.4
	3aii	4	108		64	59.3	1	101	93.5	0.47	11.8	0.62	15.5
	3b	7	108		44	40.7	1	93	86.1	0.78	11.1	0.81	11.6
	3c	15	108	0.854	35	32.4	2	83	76.9	1.43	9.5	1.47	9.8
Total for	Section A	60	173	0.847	16	9.2	6	156	90.2	3.10	5.2	2.42	4.0
B	5a	4	173		111	64.2	1	167	96.5	0.40	10.0	0.58	14.5
	5b	4	173		85	49.1	1	158	91.3	0.60	14.9	0.65	16.1
	5c	7	173		69	39.9	1	147	85.0	0.80	11.4	0.81	11.6
	5d	15	173	0.702	35	20.2	2	128	74.0	1.79	11.9	1.46	9.7
	7ai	4	85		45	52.9	1	81	95.3	0.52	12.9	0.59	14.7
	7aii	6	85		27	31.8	1	66	77.6	0.95	15.9	0.83	13.8
	7b	5	85		36	42.4	1	76	89.4	0.68	13.6	0.66	13.2
	7c	15	85	0.691	19	22.4	2	67	78.8	1.56	10.4	1.38	9.2
	8ai	4	88		48	54.5	1	84	95.5	0.51	12.8	0.63	15.6
	8aii	6	88		29	33.0	1	71	80.7	0.93	15.5	0.87	14.5
	8b	5	88		36	40.9	1	71	80.7	0.84	16.8	0.87	17.4
	8c	15	88		11	12.5	2	68	77.3	1.64	10.9	1.16	7.7
Total for	Section B	60	173	0.797	23	13.3	6	153	88.4	3.19	5.3	2.54	4.2
Total for	Script	120	173	0.881	9	5.2	12	167	96.5	4.45	3.7	3.4	2.8
							6	136	78.6				

APPENDIX C

Table C1: Repeated measures ANOVA summary tables for ETRs from Section A of GEOG1 comparing prime marks from live marking with re-marks from the re-marking exercise whilst controlling for the effects of status of prime marker and re-marker

Tests of within-subjects contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
sectionaessaymks	17.097	1.00	17.097	7.896	.005
sectionaessaymks * rmExaminerKey	17.369	2.00	8.684	4.011	.019
sectionaessaymks * pmkrstatus	1.260	1.00	1.260	.582	.446
sectionaessaymks * rmExaminerKey * pmkrstatus	2.585	2.00	1.292	.597	.551
Error(sectionaessaymks)	736.179	340.00	2.165		

Tests of between-subjects contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	8224.831	1.00	8224.831	587.571	.000
rmExaminerKey	90.650	2.00	45.325	3.238	.040
pmkrstatus	15.668	1.00	15.668	1.119	.291
rmExaminerKey * pmkrstatus	31.528	2.00	15.764	1.126	.325
Error	4759.326	340.00	13.998		

Table C2: Repeated measures ANOVA summary tables for ETRs from Section B of GEOG1 comparing prime marks from live marking with re-marks from the re-marking exercise whilst controlling for the effects of status of prime marker and re-marker

Tests of within-subjects contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
sectionbessaymks	2.874	1.00	2.874	1.332	.249
sectionbessaymks * rmExaminerKey	39.881	2.00	19.940	9.244	.000
sectionbessaymks * pmkrstatus	1.538	1.00	1.538	.713	.399
sectionbessaymks * rmExaminerKey * pmkrstatus	10.219	2.00	5.109	2.369	.095
Error(sectionbessaymks)	733.422	340.00	2.157		

Test of between-subjects effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	5515.112	1.00	5515.112	432.047	.000
rmExaminerKey	78.510	2.00	39.255	3.075	.047
pmkrstatus	5.896	1.00	5.896	.462	.497
rmExaminerKey * pmkrstatus	39.264	2.00	19.632	1.538	.216
Error	4340.125	340.00	12.765		

Table C3: Repeated measures ANOVA summary tables for whole scripts from GEOG1 comparing prime marks from live marking with re-marks from the re-marking exercise whilst controlling for the effects of number of prime markers and route through the question paper

Tests of within-subjects contrasts

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
marks	110.492	1	110.492	7.878	.006
marks * numprimemkrs	31.975	4	7.994	.570	.685
marks * route	116.178	3	38.726	2.761	.044
marks * numprimemkrs * route	193.038	12	16.086	1.147	.327
Error(marks)	2145.782	153	14.025		

Tests of between-subjects effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	567833.585	1	567833.585	2586.527	.000
numprimemkrs	1256.045	4	314.011	1.430	.227
route	615.970	3	205.323	.935	.425
numprimemkrs* route	3914.439	12	326.203	1.486	.135
Error	33588.871	153	219.535		

APPENDIX D

Question paper and mark scheme for GEOG1, January 2010

Centre Number						Candidate Number				
Surname										
Other Names										
Candidate Signature										

For Examiner's Use	
Examiner's Initials	
Question	Mark
1	
2	
3	
4	
5	
6	
7	
8	
TOTAL	

General Certificate of Education
Advanced Subsidiary Examination
January 2010

Geography GEOG1

Unit 1 Physical and Human Geography

Friday 15 January 2010 9.00 am to 11.00 am

You will need no other materials.
You may use a calculator.

Time allowed
- 2 hours

Instructions
- Use black ink or black ball-point pen. Use pencil only for drawing.
- Fill in the boxes at the top of this page.
- You must answer the questions in the spaces provided. Answers written in margins or on blank pages will not be marked.
- Answer Question 1 and **one other** from **Section A** and Question 5 and **one other** from **Section B**.
- Do all rough work in this book. Cross through any work you do not want to be marked.

Information
- The marks for questions are shown in brackets.
- The maximum mark for this paper is 120.
- Each question is worth 30 marks.
- You are expected to use a calculator where appropriate.
- You will be marked on your ability to:
  – use good English
  – organise information clearly
  – use specialist vocabulary where appropriate.

Advice
- Where appropriate, sketch maps and diagrams should be used to illustrate answers and reference made to examples and case studies.
- You are advised to spend about 60 minutes on Section A and about 60 minutes on Section B.

SECTION A

Answer **Question 1** and **one other** question from this section.

---

1 RIVERS, FLOODS AND MANAGEMENT

1 (a) Describe how water reaches a river channel in a drainage basin.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

1 (b) Shawbury is situated in the drainage basin of the River Severn.
Information on rainfall for two separate periods is shown in **Figures 1a** and **1b**.

Figure 1a

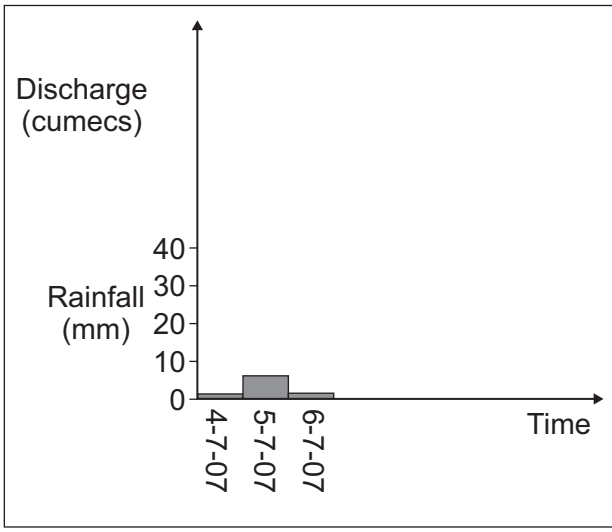
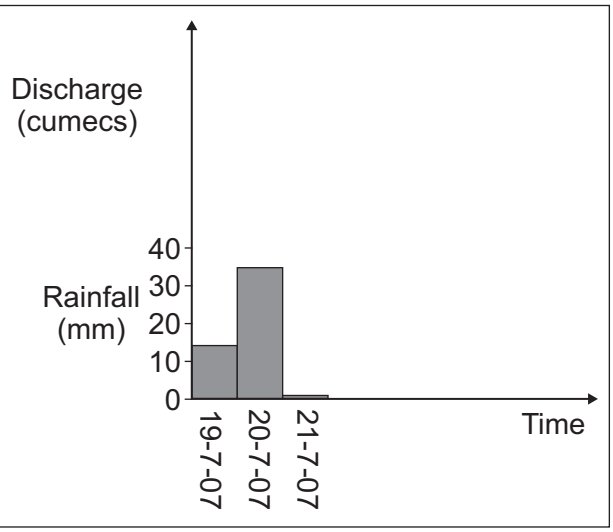


Figure 1b



1 (b) (i) Draw sketch hydrographs on **Figures 1a** and **1b** to show the contrasting effects of **each of the two** rainfall periods on river discharge.

(3 marks)

1 (b) (ii) Suggest reasons for the differences in the shapes of the hydrographs that you have drawn in 1(b)(i).

.....

.....

.....

.....

.....

.....

(3 marks)

(Extra space) .....

.....

.....

.....

.....

.....

Question 1 continues on the next page

1 (c) Study **Figure 2** which shows an area of deciduous forest.

Figure 2

1 (c) Explain likely effects of the forest on river discharge.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(5 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 1 continues on the next page

1 (d) Discuss the advantages and disadvantages of hard engineering as a flood management strategy.

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 1 continues on the next page

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

30

2 COLD ENVIRONMENTS

2 (a) Outline the global distribution of alpine cold environments.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

Question 2 continues on the next page

2 (b) Study **Figure 3** which shows a landscape formed by glacial erosion in Glacier National Park, USA.

Figure 3

2 (b) (i) Describe the landforms resulting from glacial erosion shown in **Figure 3**.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

2 (b) (ii) Choose **one** of the landforms of glacial erosion shown in **Figure 3**.

Name the landform and explain its formation.

Name of landform .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(7 marks)

Question 2 continues on the next page

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

2 (c) Study **Figure 4** which is an extract from a news report on the use of a cold environment.

Figure 4

Save our blighted Everest, says Sir Edmund Hillary

Sir Edmund Hillary, who died last Thursday, had become concerned in recent years with the piles of rubbish left by climbing parties on Everest – estimated to weigh as much as 200 tonnes – as well as the risks presented to the native population by vast lakes formed by melting glaciers as a result of climate change.

The glacier where Sir Edmund and Tensing Norgay pitched their base camp before their assault on the 8848 metre summit has retreated 4.8 km in the past 20 years. The rate of glacial melt accelerated from an average of 40 metres a year in the 40 years to 2001 to 74 metres a year in 2006.

The Himalayan Trust has applauded recent efforts to remove much of the rubbish, including large amounts of empty oxygen bottles, which litter the trail to the summit. But the creeping commercialisation of the pristine mountain which Sir Edmund would have seen continues unabated.

Some 1700 people are estimated to have reached the summit since Sir Edmund and Norgay. Growing numbers of people are willing to pay an average of £27 000 to join one of dozens of expeditions each year. At one point in 2007, more than 100 people reached the summit in a single day.

2 (c) Using **Figure 4** and your own knowledge, explain why many cold environments are fragile and how they can be cared for to ensure sustainability.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 2 continues on the next page

Turn over ▶

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

30

3 COASTAL ENVIRONMENTS

3 (a) Study **Figures 5a** and **5b** which show part of Happisburgh, Norfolk, in 2002 and 2007 respectively.

Figure 5a

Figure 5b

3 (a) (i) Outline evidence that suggests marine erosion is occurring along this coast.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

Question 3 continues on the next page

3 (a) (ii) **Figure 5a** shows revetments.

Suggest reasons why revetments were ineffective in protecting this coast.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

3 (b) Describe processes of transportation on a beach.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(7 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 3 continues on the next page

.....

.....

.....

.....

3 (c) Explain the causes of sea level change and the formation of resultant coastal landforms.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 3 continues on the next page

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

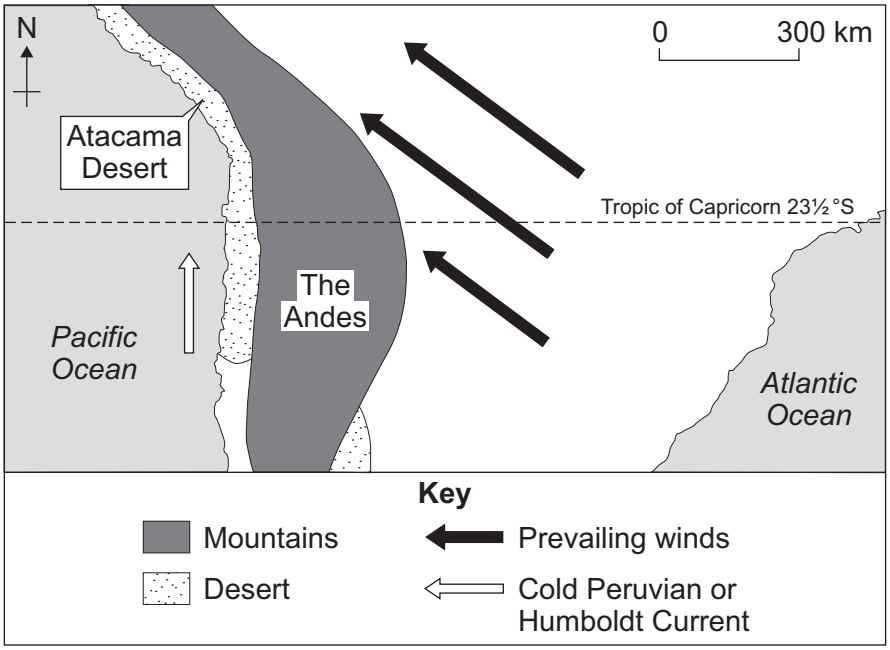
.....

30

4 HOT DESERT ENVIRONMENTS AND THEIR MARGINS

4 (a) Study **Figure 6** which shows some causes of aridity in the Atacama Desert, South America.

Figure 6



4 (a) Outline how the cold ocean current and the mountains are responsible for aridity in the Atacama Desert.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

Question 4 continues on the next page

.....

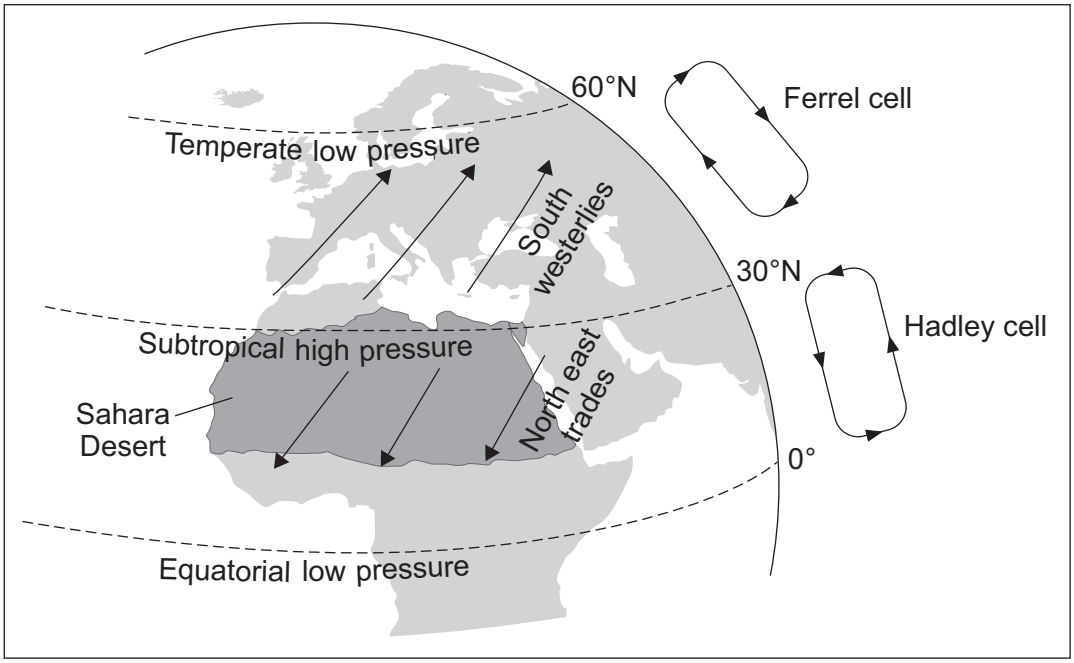
.....

.....

.....

4 (b) Study **Figure 7** which shows part of the pattern of atmospheric circulation in the northern hemisphere, and the location of the Sahara Desert.

Figure 7



4 (b) Outline how the atmospheric circulation is responsible for the aridity in the Sahara.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

4 (c) Describe the role of wind action in the formation of yardangs and zeugen.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(7 marks)

Question 4 continues on the next page

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

4 (d) To what extent do you agree that the human causes of desertification are more important than the physical causes?

Question 4 continues on the next page

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

30

Turn over for Section B

SECTION B

Answer **Question 5** and **one other** question from this section.

---

5 **POPULATION CHANGE**

5 (a) Distinguish between birth rate and fertility rate.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

5 (b) Study **Figure 8** which gives information on infant mortality rate and GDP per capita for selected countries in 2005.

Figure 8

Country	Infant mortality rate (per 1000 live births per year)	GDP per capita (US$)
Botswana	87	12 387
Brazil	31	8 402
China	23	6 757
India	56	3 452
Japan	3	31 267
Malaysia	10	10 882
Sudan	62	2 083
Thailand	18	8 677
United Kingdom	5	33 238
United States	6	41 890

5 (b) Outline the usefulness of infant mortality rate as an indicator of development.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

Question 5 continues on the next page

5 (c) Study **Figure 9** which shows average life expectancy by country in 2007.

We are unable to reproduce this figure due to third-party copyright constraints.

5 (c) Describe and suggest reasons for the pattern shown in **Figure 9**.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(7 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

5 (d) For any **two** of the following types of area, summarise the contrasts between them and explain the implications of these contrasts for social welfare:

- inner city
- suburban
- rural–urban fringe
- rural settlement.

Chosen areas .....

.....

.....

.....

.....

.....

.....

Question 5 continues on the next page

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

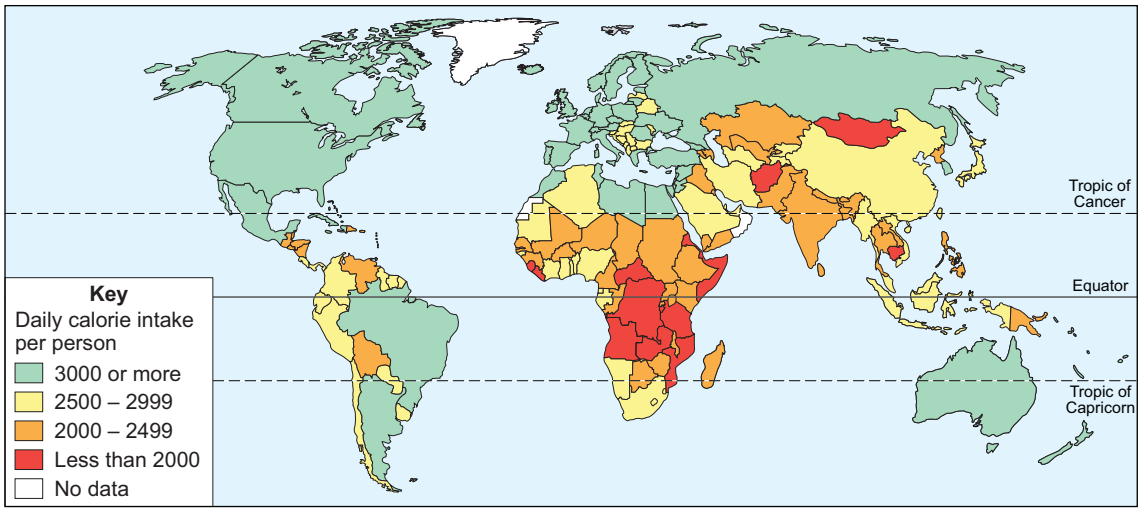
.....

Turn over ▶

6 FOOD SUPPLY ISSUES

6 (a) Study **Figure 10** which shows the average daily calorie intake per person by country in 2004.

Figure 10



6 (a) (i) Describe the pattern shown in **Figure 10**.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

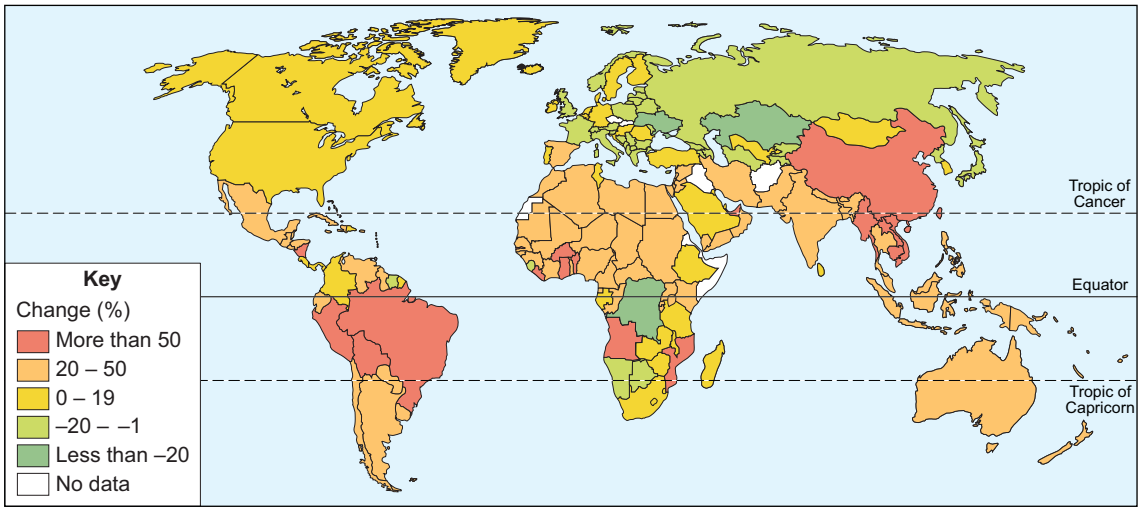
.....

.....

.....

6 (a) (ii) Study **Figure 11** which shows change in agricultural production per person from 1993 to 2003.

Figure 11



6 (a) (ii) Analyse the relationship between **Figure 10** and **Figure 11**.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(6 marks)

Question 6 continues on the next page

Turn over ▶

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

6 (b) Describe the characteristics of an intensive subsistence agricultural system.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(5 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 6 continues on the next page

6 (c) Outline strategies adopted to control food production in the EU and comment on their effectiveness.

.....

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 6 continues on the next page

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

30

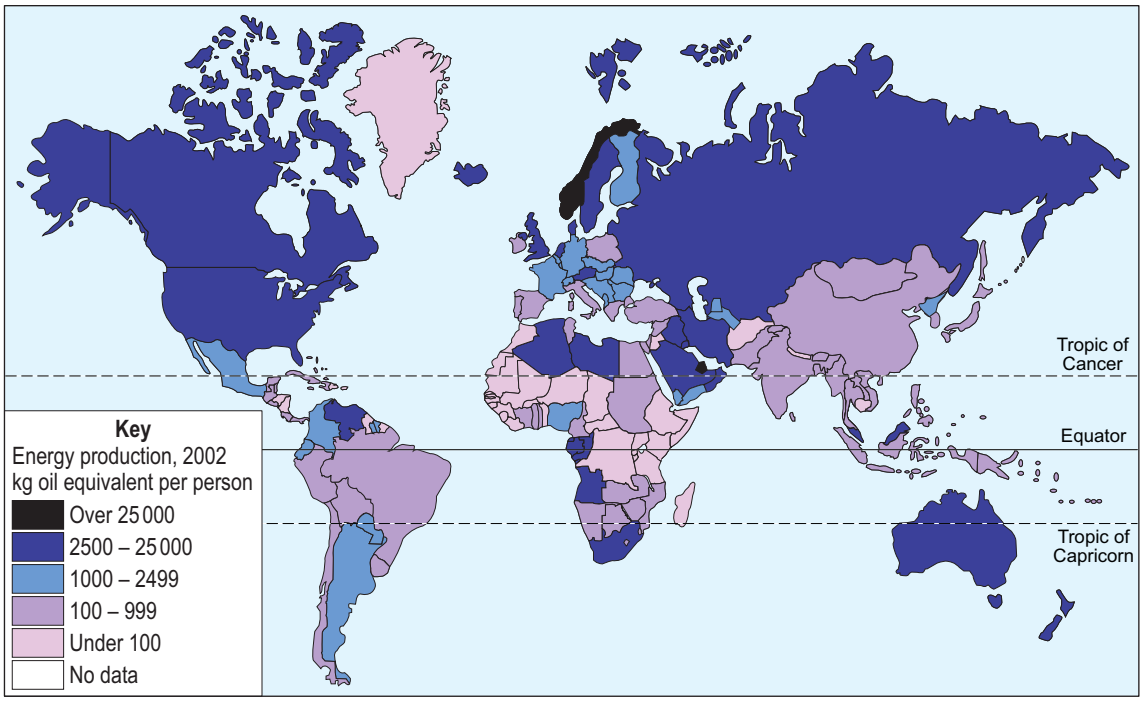
Turn over for the next question

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

7 ENERGY ISSUES

7 (a) Study **Figure 12** which shows world energy production by country in 2002.

Figure 12



7 (a) (i) Describe the pattern shown in **Figure 12**.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

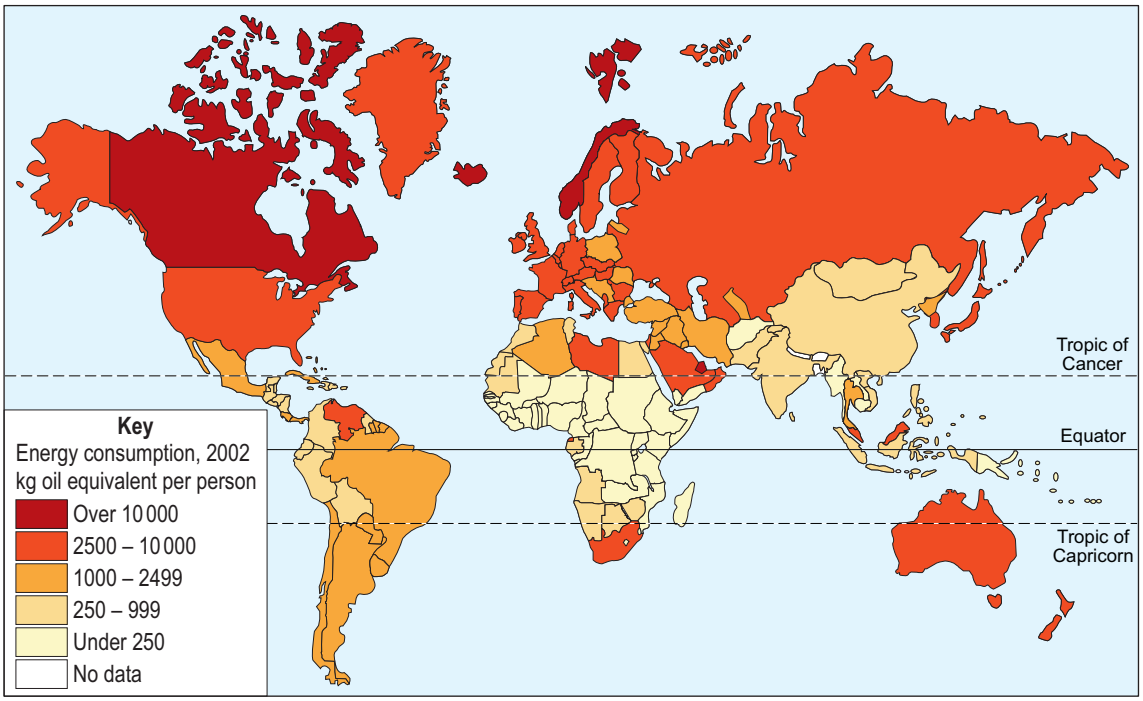
.....

.....

.....

7 (a) (ii) Study **Figure 13** which shows world energy consumption in 2002.

Figure 13



7 (a) (ii) Analyse the relationship between **Figure 12** and **Figure 13**.

.....

.....

.....

.....

.....

.....

.....

.....

Question 7 continues on the next page

.....

.....

.....

.....

(6 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

7 (b) Outline the advantages of renewable energy resources.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(5 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 7 continues on the next page

7 (c) To what extent is the global supply of energy dependent on co-operation between different countries?

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 7 continues on the next page

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

30

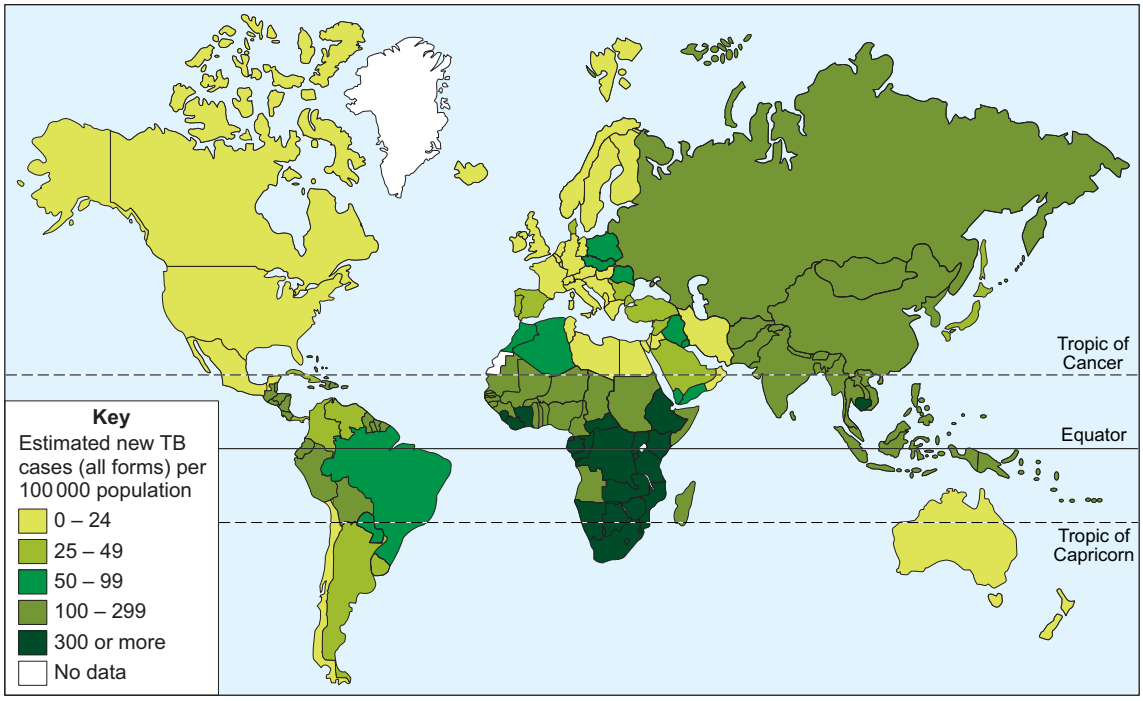
Turn over for the next question

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

8 HEALTH ISSUES

8 (a) Study **Figure 14** which shows the estimated occurrence of the infectious disease tuberculosis (TB) by country in 2005.

Figure 14



8 (a) (i) Describe the pattern shown in **Figure 14**.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

(Extra space) .....

.....

.....

.....

.....

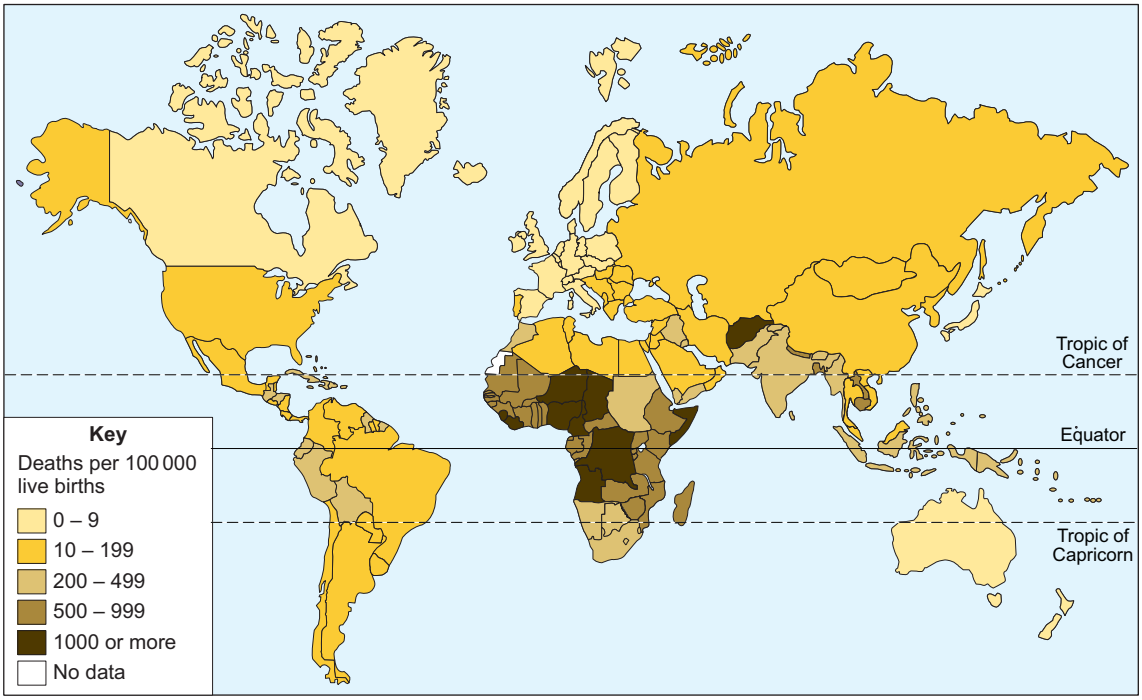
.....

.....

.....

8 (a) (ii) Study **Figure 15** which shows the global maternal mortality rate (the number of deaths of mothers per 100 000 live births) in 2005.

Figure 15



8 (a) (ii) Examine the relationship between **Figure 14** and **Figure 15**.

.....

.....

.....

.....

.....

.....

Question 8 continues on the next page

.....

.....

.....

.....

.....

.....

(6 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

8 (b) Outline the impact of **one** infectious disease on the lifestyle of the people affected.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(5 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 8 continues on the next page

8 (c) Discuss the impact of obesity on people's health and the strategies adopted to care for people with obesity.

.....

.....

.....

.....

.....

(15 marks)

(Extra space) .....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Question 8 continues on the next page

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

30

END OF QUESTIONS

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

There are no questions printed on this page

DO NOT WRITE ON THIS PAGE
ANSWER IN THE SPACES PROVIDED

ACKNOWLEDGEMENT OF COPYRIGHT-HOLDERS AND PUBLISHERS

Permission to reproduce all copyright material has been applied for. In some cases, efforts to contact copyright-holders have been unsuccessful and AQA will be happy to rectify any omissions of acknowledgements in future papers if notified.

Question 2	Figure 3:	www.shutterstock.com
Question 2	Figure 4:	Cahal Milmo, *The Independent on Sunday*, 12 January 2008
Question 3	Figures 5a and 5b:	www.mike-page.co.uk
Question 4	Figures 6 and 7:	*Arid and Semi-Arid Environments*, MICHAEL HILL, Hodder Murray, 2002, © 2002 Michael Hill. Reproduced by permission of John Murray (Publishers) Ltd.
Question 5	Figure 8:	*Human Development Report 2007/2008*, 2007, Palgrave Macmillan. Reproduced with permission of Palgrave Macmillan.
Question 5	Figure 9:	The Knowers Ark Global Education Project www.theglobaleducationproject.org
Question 6	Figure 10:	*Longman Student Atlas*, Pearson Education Ltd, 2005
Question 6	Figure 11:	A BARKER, D REDFERN, M SKINNER, *AQA AS Geography*, Philip Allan Updates, 2008. Reproduced by permission of the publisher.
Question 7	Figures 12 and 13:	from *Oxford Student Atlas*, edited by Patrick Wiegand (OUP, 2005), copyright © Oxford University Press 2005, reprinted by permission of Oxford University Press.
Question 8	Figure 14:	World Resources Institute
Question 8	Figure 15:	www.who.int/research/WHO_maternal_mortality_ratio.pdf

Version : 12/05/2010

ASSESSMENT and QUALIFICATIONS ALLIANCE

General Certificate of Education

Geography 2030

Specification

GEOG1 Physical and Human Geography

Post-Stand

Mark Scheme

2010 examination - January series

Mark schemes are prepared by the Principal Examiner and considered, together with the relevant questions, by a panel of subject teachers. This mark scheme includes any amendments made at the standardisation meeting attended by all examiners and is the scheme which was used by them in this examination. The standardisation meeting ensures that the mark scheme covers the candidates' responses to questions and that every examiner understands and applies it in the same correct way. As preparation for the standardisation meeting each examiner analyses a number of candidates' scripts: alternative answers not already covered by the mark scheme are discussed at the meeting and legislated for. If, after this meeting, examiners encounter unusual answers which have not been discussed at the meeting they are required to refer these to the Principal Examiner.

It must be stressed that a mark scheme is a working document, in many cases further developed and expanded on the basis of candidates' reactions to a particular paper. Assumptions about future mark schemes on the basis of one year's document should be avoided; whilst the guiding principles of assessment remain constant, details will change, depending on the content of a particular examination paper.

Set and published by the Assessment and Qualifications Alliance.

The Assessment and Qualifications Alliance (AQA) is a company limited by guarantee registered in England and Wales (company number 3644723) and a registered charity (registered charity number 1073334).
Registered address: AQA, Devas Street, Manchester M15 6EX *Dr Michael Cresswell Director General*

GEOG1, GEO4A and GEO4B General Guidance for GCE Geography Assistant Examiners

Marking – the philosophy
Marking should be positive rather than negative.

Mark schemes – layout and style
The mark scheme for each question will have the following format:
a) Notes for answers (nfa) – exemplars of the material that might be offered by candidates
b) Mark scheme containing advice on the awarding of credit and levels indicators.

Point marking and Levels marking
a) Questions with a mark range of 1-4 marks will be point marked.
b) Levels will be used for all questions with a tariff of 5 marks and over.
c) Two levels only for questions with a tariff of 5 to 8 marks.
d) Three levels to be used for questions of 9 to 15 marks.

Levels Marking – General Criteria
Everyone involved in the levels marking process (examiners, teachers, students) should understand the criteria for moving from one level to the next – the "triggers". The following general criteria are designed to assist all involved in determining into which band the quality of response should be placed. It is anticipated that candidates' performances under the various elements will be broadly inter-related. Further development of these principles will be discussed during Standardisation meetings. In broad terms the levels will operate as follows:

Level 1: attempts the question to some extent (basic)
An answer at this level is likely to:
- display a basic understanding of the topic
- make one or two points without support of appropriate exemplification or application of principle
- demonstrate a simplistic style of writing perhaps lacking close relation to the terms of the question and unlikely to communicate complexity of subject matter
- lack organisation, relevance and specialist vocabulary
- demonstrate deficiencies in legibility, spelling, grammar and punctuation which detract from the clarity of meaning.

Level 2: answers the question (well/clearly)
An answer at this level is likely to:
- display a clear understanding of the topic
- make one or two points with support of appropriate exemplification and/or application of principle
- give a number of characteristics, reasons, attitudes ("more than one") where the question requires it
- provide detailed use of case studies
- give responses to more than one command e.g. "describe and explain.."
- demonstrate a style of writing which matches the requirements of the question and acknowledges the potential complexity of the subject matter
- demonstrate relevance and coherence with appropriate use of specialist vocabulary
- demonstrate legibility of text, and qualities of spelling, grammar and punctuation which do not detract from the clarity of meaning.

Level 3: answers the question very well (detailed)

An answer at this level is likely to:

- display a detailed understanding of the topic
- make several points with support of appropriate exemplification and/or application of principle
- give a wide range of characteristics, reasons, attitudes, etc.
- provide highly detailed accounts of a range of case studies
- respond well to more than one command
- demonstrate evaluation, assessment and synthesis throughout
- demonstrate a sophisticated style of writing incorporating measured and qualified explanation and comment as required by the question and reflecting awareness of the complexity of subject matter and incompleteness/ tentativeness of explanation
- demonstrate a clear sense of purpose so that the responses are seen to closely relate to the requirements of the question with confident use of specialist vocabulary
- demonstrate legibility of text, and qualities of spelling, grammar and punctuation which contribute to complete clarity of meaning.

CMI+ annotations

- The annotation tool will be available for all items.
- Where an answer is marked using a levels response scheme the examiner should annotate the script with 'L1', 'L2' or 'L3 at the point where that level has been reached. At each point where the answer reaches that level the appropriate levels indicator should be given. In addition examiners may want to indicate strong material by annotating the script as "Good Level…". Further commentary may also be given at the end of the answer. Where an answer fails to achieve Level 1 zero marks should be given.
- For point marked questions where no credit-worthy points are made, zero marks should be given.
- The following is a list of the annotations available on the CMI+ system:

✓	- creditworthy point	f	- fragility
aa	- analytical approach	g	- global energy supply
adv.	- advantage(s)	hc	- human cause
c	- comment	i	- impact
con.	- contrasts	l	- link(s)
co-op.	- co-operation	o	- outline
desc.	- description	pat.	- pattern
dis.	- disadvantage(s)	pc	- physical cause
ex.	- examines/examination	s	- strategy/ies
exp.	- explanation	sus.	- sustainability
e-l	- explanation of landform(s)	twe	- to what extent
e-s	- explanation of sea level change		

Other mechanics of marking

- Examiners may add other comments or abbreviations as appropriate such as: 'rep' (repeated material), 'va' (vague), 'NAQ' (not answering question), 'seen', etc.
- Unless indicated otherwise, always mark text before marking maps and diagrams. Do not give double credit for the same point in text and diagrams.

1 (a)	2 correct terms = 1 mark. Direct precipitation/channel catch over the river channel (1). Overland flow/surface runoff where water is running over the land surface into the channel (1). Throughflow where water having infiltrated into the soil (1) flows parallel to the surface and into streams (1) above water table/unsaturated zone. Groundwater flow where water that has percolated deeper down into the rock (1) flows parallel to the surface and enters the stream through hydrostatic pressure (1) flows within saturated zone (1). Pipeflow where water follows roots of vegetation (1). May refer to the continuous nature of some of these and relative permanence (1). 4 x 1; maximum 2 marks on any one process.	***(4 marks)***
1 (b)(i)	Sketch hydrograph for 4/5/6.7.07 should show much lower peak (1) and longer time lag (1) in contrast to 19/20/21. 2×1 for each element. Peak should be much higher (approx 3 times) for third mark or much more steeply rising limb for 19 – 21. If dissimilar start point on Y axis/or does not begin at Y axis – maximum 2.	***(3 marks)***
1 (b)(ii)	High rainfall total over the two days of 19/20 suggests high intensity which would lead to a steep rising limb (1). The higher amount on the first day would reduce lag time and increase the peak as stores would be fuller (1) saturation or lack of infiltration (1). Rainfall would flow overland (1) so faster transfer (1). Must relate back to hydrographs and make links.	***(3 marks)***
1 (c)	The presence of vegetation will slow down the rate at which the water reaches the river, thus increasing the lag-time and reducing peak. This is due to the leaves especially that intercept the rainfall en route to the ground. This delays the progress of the water. Some will reach the ground as stemflow and drip, whilst some may never reach the ground and the river due to evaporation. As the forest is deciduous, the effectiveness of interception will vary, being less apparent during the winter when the leaves have been lost. The trees will also use some of the water to maintain growth and slow the speed at which water gets to the river and so the amount via overland flow and throughflow is much less. Reference to leaf litter soaking up water before infiltration. **Level 1 (1 – 3 marks)** *Outlines the effect of reducing discharge / slowing down response.* *Begins to explain.* *Some use of appropriate terminology present at the higher end.* **Level 2 (4 – 5 marks)** *Explanation is clear / sequence given.* *More than one effect needed for L2.* *Refers to the features on photograph, e.g. density of leaf cover.* *Appropriate geographical terminology is used.*	***(5 marks)***

1 (d)	A definition of hard engineering is likely to form part of the answer. This is where structures are added so that the river channel is directly interfered with so that speed of flow is altered or level of storage is changed. There is often no concern for the environment and the element of control is strong. Specification refers to dams, straightening, building up levées and diversion spillways so reference to some of these is to be expected. **Dams** e.g. Three Gorges Dam. *Advantages* – effective at regulating the flow and controlling flooding. Can have an impact on the entire river – depending on size. *Disadvantages* – schemes are costly; interfere with river processes – deposition encouraged in calm waters behind dam and clearwater erosion after the dam. Reduction in abrasion conversely. Much less discharge downstream. Impact on habitats. Displacement of potentially very large numbers of people. **Straightening** e.g. Severn. *Advantages* – as resulting route is shorter it will remove water from area faster; cheaper than dams and quicker to implement. *Disadvantages* – can cause problems downstream and exacerbate the flood risk there; can interfere with river processes – faster flow increasing erosion; impact on meander development and can damage habitats. **Levées** e.g. Mississippi. *Advantages* – can increase the capacity of river significantly and so effectively reduce flooding in certain areas. *Disadvantages* – can result in more severe flooding if levees are breached; water cannot return to river channel prolonging flood event, flood plain cannot develop. **Diversion spillways** e.g. Jubilee River, Maidenhead / Windsor area on Thames. *Advantages* – increases the capacity of the river by providing an alternative additional channel, especially during times of high flow; if done appropriately can be sympathetic to the environment and enhance it. *Disadvantages* – significant cost needed for major schemes; can lead to problems downstream and increase the flood risk.	***(15 marks)***

	Level 1 (1 – 6 marks) *Defines hard engineering.* *Describes how floods may be managed using hard engineering.* *Information likely to be generic – Basic advantages/disadvantages.* *Some use of appropriate terminology present at the higher end.* **Level 2 (7 – 12 marks)** *Describes hard engineering strategies.* *Description of how floods may be managed using hard engineering is more specific and precise.* *Clear link to advantages and/or disadvantages. (may be general references to advantages of dams). Probable imbalance to disadvantages.* *May relate to case studies.* *Appropriate geographical terminology is used.* *Begins to discuss.* **Level 3 (13 – 15 marks)** *Precise description of strategies linked to both advantages and disadvantages. (Greater balance) Focus on flooding.* *Case studies likely to be used in support.* *Specific terminology is used throughout.* *Purposeful discussion.*	

2 (a)	2 ranges named = 1 mark. High altitude/mountainous areas/above tree line (1). Most alpine cold environments are found in the northern hemisphere (1). Thus it includes areas on west coast of USA (Rockies) and east coast (Appalachians and Laurentian Mts) (1). There are smaller and more intermittent areas along the west coast of Scandinavia (1). Outliers occur in central Europe/Alps (1). Area furthest south is about 30 degrees north in Asia/Himalayas (1). The extent in the southern hemisphere is more limited with the west coast of South America showing more extensive area/Andes (1) and a part of South Island, New Zealand (1). Named location in wider context e.g. Alps in France, Italy, Switzerland (1). 4 x 1	***(4 marks)***
2 (b)(i)	Answer should relate to the landforms visible in the photograph. Question asks for description of landforms and this rather than their identification is what must be rewarded. Corries, arêtes, pyramidal peaks and truncated spur, forming steep valley side to glacial trough are visible. Also hanging valleys. Statements should relate to characteristics of these as shown on the photograph. e.g. for corrie – steep backwall is visible, corries appear to have snow/ice/glaciers in them. Arêtes appear as narrow ridges, where they meet pyramidal peaks are clear – horn shape; highest points. Max 2 for generic responses without reference to photo. More than one landform needed for 4. Need valid qualification of landform for mark (not just identification). 4 x 1	***(4 marks)***

2 (b)(ii)	The landforms that may be used are the corries, arêtes, pyramidal peaks, glacial trough, hanging valley. Responses should refer to the process of ice formation and intermediate stages where firn/neve is formed. Freeze-thaw weathering particularly on the backwall results in loose material that finds its way into the glacier. The process of plucking from back, sides and base also creates loose material within corrie glacier. This moraine is used in abrasion. The rotational slip movement creates the armchair shape of the corrie allowing a lake to occupy it after glaciation. Corries form best on north east/east facing slopes in northern hemisphere due to aspect. Expect any of processes to be described more fully. Arêtes necessitate two corries on opposite sides – role of freeze-thaw weathering may be explored further, whilst 3 or 4 are required for the arêtes to coalesce to form the pyramidal peak. Glacial trough formed where ice from corries coalesces and occupies former main river valley. Ice removes interlocking spurs to create truncated spurs of side and deepens and widens the valley via abrasion, plucking and bulldozing. Role of freeze thaw weathering, providing material within glacier is relevant. **Level 1 (1 – 4 marks)** *Begins to explain.* *Answer may be partial – and an emphasis placed on one element.* *Sequence will be incomplete.* *Some use of appropriate terminology present at the higher end.* Nivation hollow and process top L1 maximum. **Level 2 (5 – 7 marks)** *Landform is more fully explained – specific reference to processes where they operate.* *Explanation is clear.* *Sequence given so that resulting landform is clear.* *Appropriate geographical terminology is used.*	***(7 marks)***

2 (c)	**Fragile environment** is likely to be defined – an area that is susceptible to damage and one where the impact will be long-lasting due to the inability of the area to recover easily due to extreme cold where things neither decay nor grow quickly; may also be seen as susceptible to real impact by temperature change – currently increase with regard to global warming. **Sustainability** should also feature – whereby the use of the area does not lead to irrevocable damage, but leaves it for future generations to experience. An element of care and stewardship therefore is present here. The key threat given in the article refers to the development of tourism. The large amount of rubbish left by tourists is a clear feature of disregard for the landscape and environment. So too is the growing number of visitors who pay large sums of money to be able to attempt the summit – 100 in a day is not what is expected in the highest area in the world. Overuse is clearly an issue in an area that is susceptible to damage – and the warming of the climate is causing the retreat of the ice cap – clearly the environment is in danger. Similar points are likely to be made with reference to Alaska and/or Antarctica in the context of tourism. Other uses can clearly be discussed – oil, fishing. The impact of oil spills may be explored and issues relating to a return to whaling, given earlier experiences that can be used to illustrate vulnerability and the need for caution. The need to care for it relates to the need to limit numbers; to look after the local population and their needs ahead of the tourists and to remove where possible evidence of people's disregard for it by clearing away rubbish, monitor/limit numbers of fish/whales taken; similarly with oil. **Level 1 (1 – 6 marks)** *Describes activities/problems.* *Fragility/sustainability defined.* *Points made are simple and random.* **Level 2 (7 – 12 marks)** *Description is more specific and precise.* *Begins to target content to purpose – seeks to explain reasons for fragility.* *Reference to figure 4 and/or own knowledge.* *Considers links to fragility and/or sustainability.* *Points are supported in places.* **Level 3 (13 – 15 marks)** *Clear, purposeful description.* *Clear explanation of reasons for fragility.* *Reference to figure 4 and own knowledge.* *An organised account that is purposeful in responding to the question.* *Clear, explicit links between activities, fragility and care applied to ensure sustainability.* *Exemplification is used to support answers.*	***(15 marks)***

3 (a)(i)	Evidence should relate to the land area – two houses gone on the left of the photo and one at bottom plus outbuildings of another property (1). Gardens are shorter (1). Road ends abruptly (1) presence of revetments (1). The shape of the coast is different with a bay-like feature being present on the left (1). The coastal protection/revetments have been washed away (1) and there appears to be more debris on the beach to the left of the photo, indicating greater activity (1). Lack of vegetation on cliff face (1). Cliff line has moved inland/retreat of coastline (1). Any valid point. 4 x 1 per basic point; 1 x 1 per developed point	***(4 marks)***
3 (a)(ii)	Very strong winds/storms (1). Ineffective as they have not been maintained (1). Made out of wood so not very strong or normal waves too powerful (1). Sections are clearly missing (1). Possible reasons why they have not been maintained such as a change in policy – managed retreat versus 'hold the line' (1) Not cost effective (1) – given area protected – as seen as worth outlay (1). May also refer to impact following loss of sections – e.g. base of cliff now exposed to waves (1); beach removed due to loss of protection in front of cliffs (1) and so cliffs erode. Mass movement on cliffs cannot be prevented by revetments.	***(4 marks)***
3 (b)	The main process is likely to refer to longshore drift. There should be reference to the two components of swash and backwash and how the angle of approach determines movement along the beach. The process is responsible for the shifting of material along the beach. There should be reference to processes such as saltation, traction and suspension. Reference to transportation by wind. **Level 1 (1 – 4 marks)** *Describes a process/es.* *Answer focuses on aspects such as swash and backwash.* *Some use of appropriate terminology present at the higher end.* **Level 2 (5 – 7 marks)** *Description is more precise and detailed.* *A broader coverage of the processes is apparent recognition that it is transport not just along the beach and longshore drift is not the only process.* *Appropriate geographical terminology is used.* *Need 2 processes to get to top of L2.* *LSD only – 5 max.*	***(7 marks)***

3 (c)	Sea level change is the result of either eustatic or isostatic change. **Eustatic** change is a global change in sea level relative to the land. These can be a fall in sea level – as occurred during glaciations or a rise in sea level as is the current situation. This is the result of water being added following temperatures warming, glaciers melting and thermal expansion as oceans warm. Current concern regarding global warming would come into this category. **Isostatic** change occurs on a local level. Again, relative change may be positive or negative. This is the result of ice melting on land masses and the loss of the additional weight causing land masses to readjust and 'bounce up'. Similarly, plate movement at subduction zones may cause the land to rise relative to the sea. Conversely, with additional weight e.g. where there are deltas being created of substantial size, the land will sink due to the additional weight. **Landforms resulting from sea level increase*/land sinking*** – fjords, rias are likely landforms; estuaries, (submerged forest) are also creditable. **Landforms resulting from sea level decrease*/land rising*** – raised beaches and fossil cliff lines are likely responses. Explanation should be given. Reference to impact of sea level change on coral reefs is valid. **Level 1 (1 – 6 marks)** *Explains a cause of sea level change and/or a landform.* *May focus on limited range – may be one-sided – either cause or landform.* *Causes and landforms separate.* *Points made are simple and random.* **Level 2 (7 – 12 marks)** *Explains cause(s) and landform(s).* *Explanation of causes is more specific and precise.* *Begins to target content to purpose – considers causes/landforms in an organised way.* *Will begin to link cause to landforms.* *Explanation of landforms is partial in sequence.* *Some reference to both aspects, although there may be imbalance*. **Level 3 (13 – 15 marks)** *Clear, purposeful explanation of causes.* *Causes are linked to resulting landforms.* *Explanation of landforms is sequential and detailed.* *Both categories are addressed in a balanced account.* *An organised account that is purposeful in responding to the question.*	***(15 marks)***

4 (a)	Figure 6 suggests two reasons for aridity – the cold current off the coast and the presence of mountains. The cold current (Humboldt/Peruvian) resulting from water being transferred from Antarctic (1). This causes air above the Pacific off the coast to cool and condensation to occur (1). As air drifts in over land, it is warmed and moisture is evaporated (1). Any winds crossing the cold water will cool, rainfall may occur over water (1); its capacity to hold moisture is limited and it too will warm up over the land (1). The Andes are the cause of a rainshadow effect (1). The air coming from the east has to rise over the mountains (1), will cool and condensation will occur (1). Thus rain will fall as the air ascends; as it descends the air will be drier and will be warmed, increasing its ability to hold moisture and so rain will not fall. 4 x 1 per basic point, 2 x (1+1) per developed point. Any combination. Maximum 3 on either mountains or cold current.	***(4 marks)***
4 (b)	The Sahara Desert is found in an area of high pressure (1). Here air that has risen near the Equator/further north in temperate low (1) has cooled/become denser (1) and so sinks to the surface (1). It is dry and its descent makes rainfall improbable (1). The North East Trades blow offshore from the continental interior (1), these have blown across large extents of land so rainfall is improbable (1). Sinking / high pressure interchangeable. 4 x 1	***(4 marks)***
4 (c)	Wind action causes yardangs – ridges of hard rock that stand above eroded lines of soft rock – etched out by prevailing wind, zeugen – ridges composed of a hard cap rock with underlying soft rock that is eroded more. Abrasion is the key process in forming both these landforms as material (sand) carried by the wind via saltation and suspension hits the masses of rock and wears it away – similar to sand blasting as is used to clean old buildings. This effect is most effective within 1.5 metres of the surface and causes rock pedestals/mushroom rocks to occur in a less extensive fashion than zeugen. The formation of such features is dependent on the geology of the area also. **Level 1 (1 – 4 marks)** *Describes the landform(s).* *Shows some awareness of what the wind does.* *Some use of appropriate terminology present at the higher end.* **Level 2 (5 – 7 marks)** *Description is more precise – some reference to both landforms.* *Aware of the role of the wind – the abrasion process and how it etches out ridges dependent on rock alignment.* *Appropriate geographical terminology is used.*	***(7 marks)***

4 (d)	Likely to begin by defining desertification – the expansion of the desert into areas that previously would not have been classified as part of it. The process is seen as being long term (unlike drought) and has serious repercussions for people as the ability of land to be productive effectively has disappeared – in a temporary or even permanent sense. **Human causes** – relate largely to increasing population and pressure exerted on the land for survival. Thus linked to this is the increased demand for fuelwood, increased pressure put on for crops, overgrazing, increased permanence of people, instead of being nomadic. The inter-relationships and inter-dependencies of these should be explored and the effects of increased pressure on the soil regarding exhaustion and erosion should be considered; as so too the loss of vegetation and effects on moisture. Poor management may be considered a cause. **Physical causes** – relate to climate change – the reduced amounts of rainfall and falls in rainfall reliability. The impact of increased temperatures on evaporation. Some may relate climate change to global warming. Response will depend on where specific aspects/areas have been studied. **Level 1 (1 – 6 marks)** *Describes some causes of desertification.* *May focus on limited range – may be one-sided and have human causes only.* *Points made are simple and random.* **Level 2 (7 – 12 marks)** *Description of causes is more specific and precise.* *Will refer to both causes, but may be clear imbalance (to human).* *Begins to target content to purpose – considers how the causal factors do cause desertification – partial sequence of events.* *Tentative/implicit assessment of 'to what extent' based on evidence.* **Level 3 (13 – 15 marks)** *Clear, purposeful description of causes – both physical and human referred to in more balanced account.* *An organised account that is purposeful in responding to the question – clear sequence between the cause and the process/concept.* *Clear, explicit assessment of 'to what extent' rooted in evidence.*	***(15 mar8s)***

5 (a)	Birth rate refers to the number of (live) births per 1000 per year (1). Fertility rate is the number of (live) births per 1000 women per year (1) aged 15 – 49 (1). It can also be defined as the number of children born on average by each woman (1), so that 2.1 would reflect the replacement rate (1). The differences relate to the group of people involved – the birth rate refers to the whole population (1), whilst the fertility rate refers to the female component only (1) and in its most specific sense is an age specific rate relating to those of child-bearing age (1). Allow up to 3 marks for definitions and 3 marks for establishing the differences. An explicit comment is needed on distinction of 4. 4 x 1	***(4 marks)***
5 (b)	Infant mortality is the number of deaths under 1 (per 1000 live births per year) (1). It reflects the deaths in one of the most vulnerable age groups (1); those who are most susceptible to die from famine, as a result of floods, earthquakes, epidemics (1). It implies much about levels of development and standard of living (1) regarding access to clean water (1), medical care/hospitals/doctors/ vaccines/drugs (1). High levels are often responsible for high levels of births to ensure the survival of most of family (1). Responses may be generic or evaluate with specific reference to data; extent of correlation may be considered or its usefulness may be questioned. 4 x 1	***(4 marks)***
5 (c)	Highest life expectancy – 77+ - is found in Japan, Australia and western Europe USA and part of southern South America -74+. Lowest life expectancy occurs in much of sub-Saharan Africa where it is below 50. It reaches 41 in significant area. Apart from the countries along the Mediterranean, life expectancy is below 64. Much of Asia is 58+ although China and parts of Malaysia, Thailand have higher levels at 69+. Reasons relate to levels of development – with generally the more developed areas having higher life expectancy. However, this is not always the case – China as an emerging world power and NIC has life expectancy of 69+, whilst the ageing population in some areas may reduce the figures, so too may lifestyle. Lowest life expectancy in Africa can be linked to the prevalence of AIDS. Generally, access to appropriate food sources, medical care etc. will explain the differences in levels globally. **Level 1 (1 – 4 marks)** *Describes the pattern generally and randomly.* *Begins to explain.* *Emphasis is likely to be on one component.* **Level 2 (5 – 7 marks)** *Description of pattern is clear (fuller, more detailed) and supported by evidence. May identify exceptions.* *A balanced account where explanation is clearly addressed.* *Appropriate terminology is used.*	***(7 marks)***

5 (d)	Content will depend on two areas selected. Whatever the areas, contrasts should be expected on the following aspects – housing – characteristics, quality; ethnicity – origin of population; age structure of population; services present; wealth of population and type of jobs people do/level of employment. These aspects are given in specification, but other alternatives are permissible if appropriate and may be substitutes – spec uses term 'such as'. The final aspect considers the implications for social welfare. This may also be integrated or candidates may choose to do this section separately. Here, there should be reference to how people's well-being is affected by where they live and the general health/well-being of the community should be considered. Thus, links between location, quality of housing and health could be considered; quality of education and achievement – number of GCSE A* - C, extent to which the area is safe, extent to which there is access to appropriate services – shops, schools, medical centres, sports facilities, meeting places, etc. The response should summarise the contrasts – these should be integral and clearly drawn out. **Level 1 (1 – 6 marks)** *Describes the characteristics and/or social welfare of the two areas. These are separate.* *General statements – applicable to any area.* *Points made are simple and random.* **Level 2 (7 – 12 marks)** *Description is more specific and precise.* *Contrasts are drawn between the two areas (maybe implicit) and the summary is clear.* *Points are supported in places.* *Begins to make links to social welfare at top end.* **Level 3 (13 – 15 marks)** *Clear, explicit purposeful summary of contrasts (maybe integrated).* *An organised account that is purposeful in responding to the question.* *Exemplification is used to support answers – case studies are effectively used.* *Clear, explicit links to social welfare and contrasts between areas*.	***(15 marks)***

6 (a)(i)	Highest levels of calorie consumption in North America, much of Western Europe and northern Asia (1), as well as Australasia, and parts of South America (1). There are some areas in northern Africa that could be viewed as exceptions to general pattern (1). Parts of eastern Europe have areas between 2500 and 2999 (1) as do parts of southern Asia (mainly China) and the north western areas of South America (1). The 2000 to 2499 category is found mainly in Africa – central, west – and parts of Asia – including India and Pakistan. The lowest consumption occurs in southern central Africa (1), with isolated occurrences in Asia, Afghanistan, Cambodia and Mongolia (1). The above represent possible statements; many other possibilities are apparent. Allow 4 x 1 for description of pattern, which should seek to establish key contrasts in calorie consumption.	***(4 marks)***
6 (a)(ii)	Answers will vary depending on which aspects of the two figures candidates focus on. For example, areas of fastest growth of agricultural production cuts across all categories of consumption – including parts of South America at highest level, China at next level of consumption and even some southern African countries that were at the lowest level of consumption – such as Angola and Mozambique. Areas that had the highest consumption generally have lower levels of increase in production, e.g. North America and Australia, whilst some have experienced a fall in production such as UK, France, Russia. Parts of central Africa, where the lowest calorie consumption was found, have experienced the highest decreases in food production. **Level 1 (1 – 4 marks)** *Describes features on Figure 11.* *Some (tentative) links made with Figure 10.* **Level 2 (5 – 6 marks)** *Relates to both Figures 10 and 11.* *Purposeful investigative approach seeks to consider links between different levels of consumption and changes in agricultural production – categorises.* *Provides evidence.* *Assesses/comments on relationship or note exceptions.*	***(6 marks)***

6 (b)	Response will depend on agricultural system selected. Likely that rice in SE Asia will be exemplar. Outline should relate to the characteristics of intensive farming where there is a high level of inputs such as labour, fertilisers, to ensure maximum output from available land. It is the input of a lot of labour where many people work to ensure high productivity. In the case of subsistence farming where the main aim of production is to supply a source of food for the farmer and family, it is the labour – often that of the extended family – that reflects the intensive aspect. **Level 1 (1 – 3 marks)** *Describes some features of either intensive and/or subsistence farming system.* *Sees aspects separately – writes generally.* **Level 2 (4 – 5 marks)** *Description is clear and purposeful.* *Sees the two aspects in a linked way.* *May relate to an actual agricultural system.*	***(5 marks)***

6 (c)	**Strategies likely to be referred to: subsidies, tariffs, intervention pricing, quotas, non-market policies, set-aside and environmental stewardship.** Initially strategies were to increase food production. Subsidies where payments are given to producer to help with costs of keeping animals etc. Tariffs were used to tax imports and make them more expensive in contrast to home-grown produce and therefore, make them less attractive. Intervention pricing gave a guaranteed return to farmers, irrespective of demand. The measures designed to increase food production were successful as shown by the butter mountains and wine lakes. Thus, the strategies changed to limit production. Quotas were introduced on certain types of production, notably the dairy industry and attempts were made to improve the quality of produce such as wine and olive oil by paying growers to reduce area of crop and to increase quality. Available subsidies were to be spread more thinly – allowing for new member states from Eastern Europe. Key policies sought to take land out of agricultural production such as set-aside where farmers were paid to leave up to 10% of land unproductive, FWS where trees could be planted as an alternative. Increasingly payments are related to concern for the environment and production being in an environmentally friendly way, rather than maximising production. Such schemes attract payments from DEFRA. **Level 1 (1 – 6 marks)** *Describes strategies used to increase or decrease food production.* *Links to food production may be implicit.* *Points made are simple and in a random sequence.* **Level 2 (7 – 12 marks)** *Begins to target information to purpose in an ordered response.* *Links strategy to food production.* *May focus on attempts to increase or decrease food production.* *Points are made with some support from EU experience.* *Tentative/implicit comment.* **Level 3 (13 – 15 marks)** *Clear, purposeful summary of strategies used to increase and/or decrease food production.* *Support is given from EU experience throughout.* *Clear, explicit comment.*	***(15 marks)***

7 (a)(i)	Highest levels of production (over 2500) in North America, parts of northern Europe (UK, Norway), Russia (1). Cluster in Middle East (1); scattered countries in Africa and only one in South America – Venezuela (1). Much of Africa comes into the lowest category – especially central areas (1) and this continent shows greatest contrasts (1). Asia has relatively low levels of production per head (1) with the exception of Malaysia (1). The above represent possible statements; many other possibilities are apparent. Allow 4 x 1 for description of pattern, which should seek to establish key contrasts in energy production.	***(4 marks)***
7 (a)(ii)	Answers will vary depending on which aspects of the two figures candidates focus on. For example, areas of highest production (top two categories) often, but not always, coincide with areas of highest consumption (top two categories) e.g. North America, much of northern Europe and Russia as well as parts of Africa – Libya, South Africa, Venezuela in South America, Australia, etc. However, some areas of highest production have lower levels of consumption, e.g. Algeria, Angola. Conversely, some areas with the highest levels of consumption have lower levels of production, e.g. much of western and southern Europe, Japan and South Korea. Africa has the lowest levels of consumption in much of the continent, reflecting its lower levels of production in the central areas especially and its lower ability to pay. **Level 1 (1 – 4 marks)** *Describes features on Figure 13.* *Some (tentative) links made with Figure 12.* **Level 2 (5 – 6 marks)** *Relates to both Figures 12 and 13.* *Purposeful investigative approach seeks to consider links between different levels of consumption and production – categorises.* *Provides evidence.* *Accesses/comments on relationship or note exceptions.*	***(6 marks)***

7 (b)	Advantages likely to include reference to the fact that they will not run out and therefore will always be available. Management may nevertheless be needed, e.g. to ensure that the use of such resources does not outstrip rate of replacement, e.g. fuelwood, biomass. Many of the resources are continually being formed – as in the case of solar, wind and tidal power. In addition, these resources are seen as clean – as they are non-polluting. Their use does not give rise to harmful by-products – carbon dioxide, etc. and cause environmental damage – acid rain and global warming may feature as may the impact on the landscape of opencast mining and spoil heaps. Advantages only required here. **Level 1 (1 – 3 marks)** *Describes features of renewable energy resources.* *Some reference to advantage(s).* **Level 2 (4 – 5 marks)** *Is aware of the advantages.* *Points are clear and developed.* *Selects and summarises key points – is clear and purposeful.*	***(5 marks)***

7 (c)	Content will depend on examples studied, but likely to include reference to: the role of trading groups – notably OPEC where prices are meant to be fixed within a range (but doesn't always work) and there is conflict between member states; role of TNC's in exploration, development in poorer countries; pressure put on by countries to increase production and reduce costs; the extent to which countries use energy resources as a source of power or where their internal politics or relationships with other parts of the world can result in supplies being threatened – such as Iran – where relations with the west are far from ideal and there has been talk of US action; Iraq – sanctions in place in days of Hussein but an uneasy peace at present having an impact on supplies; Russia – recently in news regarding gas and impact on supplies in Ukraine and return as world power, largely due to state control of oil and gas resources; May consider role of 'home-produced' resources and extent to which countries can be self-sufficient throughout world, e.g. UK, USA, China and therefore question the need for co-operation. **Level 1 (1 – 6 marks)** *Describes some of the aspects of global energy supply.* *Limited support.* *Links to question are tentative.* **Level 2 (7 – 12 marks)** *Begins to develop points and sequence them.* *Offers some support.* *Engages with theme and considers clearly one aspect relating to co-operation.* *Tentative/implicit assessment of 'to what extent' based on evidence*. **Level 3 (13 – 15 marks)** *Develops points and sequences them.* *Purposeful response with support present.* *Focus is on co-operation with necessary exemplar material.* *Clear/explicit assessment of 'to what extent' rooted in evidence.*	***(15 marks)***

8 (a)(i)	Highest incidence of cases are found in much of sub-Saharan Africa – central and southern (1) – with single outlier in Asia – Cambodia (1). Much of Asia, North Africa and western parts of tropical South America have between 100-299 (1). The next to lowest rates are found in parts of eastern Europe, Iberia, and southern areas of South America (1), whilst lowest rates are found in North America, much of western Europe, Australia (1) and three countries in North Africa – that might be seen as exceptions to the pattern (1) as might some of the Middle East countries including Saudi Arabia (higher) (1). The above represent possible statements; many other possibilities are apparent. Allow 4 x 1 for description of pattern, which should seek to establish key contrasts TB occurrence.	***(4 marks)***
8 (a)(ii)	Answers will vary depending on which aspects of the two figures candidates focus on. For example, areas of lowest incidence of TB sometimes coincide with lowest incidence of maternal death – Canada, much of Western Europe, Australia. However, some countries with lowest incidence of TB come into penultimate lowest for maternal deaths – USA, Brazil, eastern Asia. The highest incidence of maternal deaths largely coincides with areas of highest incidence of TB – areas of central and southern Africa. However, there is not a perfect match, with the maternal mortality rate being in the median category in the southern tip of south Africa and in the highest category for TB incidence. **Level 1 (1 – 4 marks)** *Describes features on Figure 15.* *Some (tentative) links made with Figure 14.* **Level 2 (5 – 6 marks)** *Relates to both Figures 14 and 15.* *Purposeful investigative approach seeks to consider links between different levels of maternal death and incidence of TB – categorises*. *Provides evidence.* *Assesses/comments on relationships or note exceptions.*	***(6 marks)***

8 (b)	Response will depend partly on disease studied – HIV/AIDS is likely to be dominant, malaria, flu probably also. Impacts are relatively generic and likely to include ill-health; reduced life expectancy; access to treatment depending on location; inability to work; dependent on other members of family/state; may be reference to discrimination, people's attitudes to certain diseases. Should be clear reference to one disease as this is a requirement of the specification. **Level 1 (1 – 3 marks)** *Describes impacts – probably in a random order.* *General points with limited support.* **Level 2 (4 – 5 marks)** *Aware of varied impacts – will categorise.* *Specific, elaborated points with support.*	***(5 marks)***

8 (c)	**Impact:** Too much food results in overweight and at its extreme, **obesity**, where the relationship between height to weight is commonly used as a measure – Body Mass Index – WHO have 25 and 30 as indicators of overweight and obesity. Impact here is likely to relate to onset of other illnesses relating to being overweight/obese – heart disease, strokes, type 2 diabetes, certain cancers, impact on hips, osteoarthritis. **Strategies to care**: **Obesity** – increased education regarding health, diets and lifestyles; encouraging access to sports centres and exercise – walking/biking to school/work; responsibility taken by marketing strategies/retailers to offer healthy options/labelling food; schools dinners initiatives; health care offer preventative checks, warnings rather than dealing with results – changing role of GPs; role of media. **Level 1 (1 – 6 marks)** *Describes the effects of obesity on people's health and/or strategies in no particular order.* *Limited support.* **Level 2 (7 – 12 marks)** *Begins to develop points of effects with reference to obesity.* *Begins to consider strategies to care for people and links to impact.* *Support is present – reference to relevant areas/strategies.* *Begins to discuss.* **Level 3 (13 – 15 marks)** *Clear, purposeful description of effects of obesity on people's health.* *A balanced answer – considers strategies to care and links clearly to impact.* *Response is precise, elaborated – support is present and targeted to the task – discursive.*	***(15 marks)***