

Research and Information Services

Monitoring Standards Report



Comparison of History Intermediate 2 2003, 2006 and 2007

Information about Monitoring Standards

Since 1998 SQA has been monitoring a sample of qualifications annually to determine whether standards in the level of demand in the syllabus and question papers are being maintained year on year and whether any changes to marking instructions have had an impact on candidates' examination results.

NQ teams are asked to compare the Arrangements, Question papers and Marking Instructions for the years in question and consider whether there have been any changes to these which have resulted in any change to the level of demand or difficulty. Teams decide whether the breadth and depth of the syllabus has remained the same and if not, how these changes have altered the level of demand in the question papers in the years being compared. The clarity of instructions and allocation of marks in the marking instructions across the years is also considered. The candidate evidence is compared and the teams decide whether or not the evidence is demonstrating the same standard of performance across the years.

NQ Monitoring teams comprise a subject specialist from Higher Education or a Professional Organisation, who is the chair of the group, the Principal Assessor for the subject, who is the SQA expert in the group and a practicing classroom teacher who can be an SQA marker.

Summary of findings

The panel, which comprised a University lecturer (chair and report writer), the Principal Assessor and a Principal Teacher concludes the overall level of demand for History Intermediate 2 has remained the same for 2007, compared to 2006 and 2003.

◆ Cohort of candidates

The number of S4 candidates for History Intermediate 2 has almost doubled since 2003 and as this year group has the highest success rate, it could be expected that the general ability of the population would be higher year on year.

◆ Syllabus

There have been no changes to the content or level of demand of the syllabus across the years.

◆ Assessment Instruments

The panel concludes that the structure of the examination paper has not changed and although questions had become homogenised in 2007, the level of demand had not altered.

◆ Marking Instructions

There was a change in 2007 in the instructions to markers of the Extended Response which provides greater clarity over how each mark relates to criteria of assessment. The panel see this as a welcome improvement. No significant change in the marking instructions for the examination scripts was identified.

◆ Scripts

The panel made a comparison of scripts representative of different awards and

concludes that across the years, the same grades show comparable performance in the small sample available.

1 Introduction

Are there any general changes, eg in teaching practice or in the population of candidates, that may have influenced the general ability of candidates in one of the years, in particular the way they meet the demands made by the syllabus and exam?

A change in the population of the candidates should be noted. The number of candidates has increased markedly, from 2,011 in 2003 to 3,980 in 2007, an increase of 98%. The most significant change, however, has been a marked increase in S4 candidates. In 2007 44% were S4 and 46% S5. In 2006 39% were S4 and 50% S5. By contrast, in 2003 only 11% were S4, and 75% were S5.

It may be inferred that S4 candidates are students who are moving straight to Intermediate 2, while those in S5 have taken Standard Grade in S4. Those in S5 taking Intermediate 2 are therefore likely to be students who have not performed sufficiently well in Standard Grade to be permitted to move directly to Higher, or who have failed a NAB at Higher and are taking Intermediate 2 as a 'fall back'. S4 candidates, by contrast, are likely to include students capable of proceeding to Higher in S5. It might be expected, therefore, that a higher proportion of S4 candidates should mean that the general ability of the population of candidates will be higher.

The better ability of S4 candidates is born out by the figures. The pass rate for S4 candidates in 2006 was 88%, but only 56% for S5 candidates; in 2007 the pass rate for S4 candidates was 93%, and 63% for S5 candidates. Equivalent figures for 2003 are not available. The higher general ability of candidates overall is not, however, so clear cut. It might be expected that the overall pass rate in 2006 would be greater in 2006 than in 2003. This is not the case: in 2006 the overall pass rate was 69%, but it was 71% in 2003. It is only in 2007 that a notable increase is apparent, with the overall pass rate rising to 77%. It may also be noted that in 2006-2007 the percentage of S3 candidates remained steady at 2% but that the pass rate rose from 68% to 90%.

Further comment on 2007 is provided in the final statement at the end of this report.

2 Level of demand of arrangements / syllabus

Was the syllabus in 2007 more/no more/less demanding than in 2006 and 2003?

The content of the syllabus has not changed, and neither have the processes relevant to the discipline of History.

Please explain using the next questions and other considerations if relevant to the level of demand.

2.1 Are there differences in breadth between the arrangements in different years and how has this changed the overall demand?

- a Was the breadth of the syllabus in 2007 more/no more/less demanding than in 2006 and 2003?

The breadth in 2007 was no more demanding than in previous years.

- b Were there additions to the more recent arrangements which were not included before?
There were no additions.
- c Has anything been removed from the more recent arrangements?
Nothing has been removed.
- d Have any parts in the arrangements been changed (in approach, content or otherwise)?
Nothing has been changed.

2.2 Are there differences in depth between the arrangements in different years and how has this changed the overall demand?

There have been no differences in depth.

Please explain using the next questions (and other considerations if relevant to the level of demand)

- a Was the amount of detail covered in the syllabus in 2007 more/no more/less demanding than in 2006 and 2003?
The amount of detail was no more demanding.
- b Was there the same or a different emphasis on integration of skills and content?
The emphasis on integration of skills and content was the same.

3 Level of demand of examination questions

Are there differences in the questions that have made them more or less demanding?

Please explain using the next issues and other considerations if relevant to the level of demand.

- a. General approach in the assessment instruments
The approach to the assessment of the Extended Response (previously Paper 2) since 2007 is more structured than in previous years, with the aim of achieving greater consistency. (See 4a below)
- b. Coverage of the syllabus
The coverage of the syllabus in exam questions has not changed.
- c. Structure of the total assessment task
The structure of the total assessment task has not changed.
- d. Demands of assessment tasks/questions
The panel noted three ways in which questions had become homogenised.

- (i) In 2007 all sources consisted only of text; in previous years at least one (2006) or as many as three (2003) pictures/cartoons had been used. It need not be inferred from this, however, that pictures will no longer be used.
- (ii) In 2007 the source extracts within each type of source question were of similar length; in previous years there was some variety.
- (iii) In 2007 all short essay questions began ‘Explain...?’ (eg Explain why...?); in 2003 there were a few which began ‘How far...?’ or ‘How successful...?’

It was not clear to the panel that homogenisation in these cases necessarily meant that, in terms of the criteria of assessment, questions had become any easier or harder. The possibility was raised that it might be more difficult to write a conclusion for a short essay question beginning ‘Explain...’ than for a question beginning ‘How far...?’ or ‘How successful...?’; the very limited evidence of the sample scripts did not offer enough material to pursue this further.

The panel also noted that there was variety in the number of different types of source comparison questions (taking the contexts as a whole). Most were of the type ‘How far do sources A and B agree/disagree?’. The number of questions of the type ‘Compare sources A and B as evidence’ or ‘Compare the views in sources A and B’ ranged from 11 in 2003, 1 in 2006, and 5 in 2007. Again, it was not clear that this necessarily meant that questions had become any easier or harder. The panel noted that there could be a perception that questions of the type ‘How far do sources A and B agree/disagree?’ were easier because the candidate’s response would be limited to the material provided in the source. This perception would be reinforced by the finding (see 5.1 and 5.2, below) that weaker candidates seemed to find recall more challenging than working solely with sources. A distinction should be drawn, however, between ‘pure’ recall (in the ‘describe’ question) and the less demanding ‘prompted’ recall in a source question.

4 Level of demand of examination marking

Are there any differences in the marking instructions that made it more or less difficult to obtain the same grade?

Please explain using the next issues and other considerations if relevant to the level of demand.

a. Different type of instructions

A significant change in 2007 in the instructions to markers of the Extended Response (previously Paper 2) was noted. Instead of describing criteria by each grade (A, B, C) in relation to overall performance, the criteria were broken down into three aspects (Introduction/Conclusion, Argument, and Knowledge and Understanding), and a descriptor given for each mark. Introduction/Conclusion had a maximum mark of 5; Argument a maximum of 6; Knowledge and Understanding a maximum of 9 (the same proportions as in Higher and Advanced Higher).

The panel felt that this innovation did *not* amount to a significant change in criteria. What was new was the way existing criteria (with one exception) were to be implemented. As such it did *not* amount to a significant change in the intended level of demand of

examination marking. (The one exception was that awareness of different possible interpretations, which had appeared in the descriptor for grade A, had been removed because this was felt to be more appropriate for Higher level.)

In two key respects, however, this does represent a change in marking practice.

- (a) Previously the appropriate balance between Introduction/Conclusion, Argument, and Knowledge and Understanding could only be gauged subjectively by markers. As of 2007, the maximum marks for each of these aspects made the balance between them unambiguous.
- (b) The descriptors for each mark for Introduction/Conclusion and for Argument made assessment more consistent as well making it easier for markers to apply the criteria consistently. It also made marking practice more transparent for checking or review. The likely outcome was greater consistency in performance by markers across the whole range of the mark-scale. (For further comment, see 5.1 below.)

The only additional sense in which this could amount to a change in examination marking, therefore, was by making it easier to bring potentially unintended marking practice into line with what was appropriate to this level. Possible examples of unintended practice that would be minimised or prevented include:

- (i) A marker who might be reluctant to use the top of the marking scale.
- (ii) A marker who might expect a level of coherence that would be more appropriate for Higher level than Intermediate 2.
- (iii) A marker who might expect an element of historiography to be present if an extended response was to be awarded a very high mark.

All these instances of unintended marking practice would result in candidates being awarded a lower mark than they deserved according to a correct reading of the marking criteria. It should be noted that the panel also found instances in the sample of marked scripts of very high marks being awarded in 2003 which would not have received such high marks if the criteria had been applied correctly. (For further comment, see 5.1 below.)

It was also noted that the introduction of 'A' for Argument to the markers' codes used on the scripts themselves would make it easier to identify and correct any inconsistent marking practice should this still occur.

b. More or fewer examples

No significant change in this aspect of the marking instructions was identified. The only point that was noted was that, among the examples supplied for markers in assessing 'describe' questions in 2003 there were instances, in a few contexts, where only a single word or brief phrase was given. The injunction that candidates should give *supported* examples was in all cases stated explicitly. In 2006 and 2007 this was reinforced by supplying markers only with supported examples, eliminating the one-word or very brief instances.

c. More or less detail

The only change that was noted was that description statements now follow exactly the same pattern. The panel did not feel that the statements in 2003 were in any way ambiguous, so this change should not have had any impact on the level of demand of marking.

5 Grading of candidates' performances

Do scripts with the same grades in 2007 2006 and 2003 show comparable performance

5.1 Are Grade A band 2 scripts from 2003 and 2006 better than/comparable to/not as good as the same grade scripts from 2007? Compare Grade C band 6 scripts for 2003 and 2006 with 2007.

Because the only significant change in the instructions to markers has been in relation to the Extended Response (previously Paper 2) in 2007 (see 4a above), the panel concentrated its attention first on comparing performances in this element in 2007 as against 2003 and 2006. In the sample of scripts all those who attained A band 2 or C band 6 had been awarded A-range marks in their Extended Response/Paper 2. The panel therefore began by examining examples of Extended Essay/Paper 2 from 2007, 2006 and 2003 that had been awarded 18, and then spread their net wider to include examples of 16 and 14. The panel then turned its attention to the Exam (Paper 1), beginning with examples of those who had achieved 25 in 2007, 2006 and 2003 (who, due to outstanding Extended Responses/Paper 2, had ended up at or near A band 2), and those who had achieved 19 in 2007, 2006 and 2003 (who, due to A-range Extended Responses/Paper 2, had ended up with C band 6). The panel finished by looking at the scripts of candidates in 2007, 2006 and 2003 who had attained 31-33 in the exam.

Extended Response (Paper 2)

The panel found some indication of inconsistency in the way marks high in the A-range had been applied in 2003 (but not in 2006), and took the view that this would not have occurred with the new instructions implemented in 2007.

The panel also found an instance of an Extended Response in 2007 which was (correctly) awarded 14, but which could have been marked as a B according to the criteria given to markers in 2003 and 2006 because the balance between aspects (Introduction/Conclusion, Argument, and Knowledge and Understanding) was ambiguous.

It should be noted that in nearly all cases the discrepancy was within the accepted level of tolerance of one mark, and no instances were noted where a candidate's overall award would have changed.

In all years it is apparent that candidates who achieve less than 'A' are brought down by their performance in the exam, with the Extended Response (previously Paper 2) as their better mark. The pattern of performance in the sample scripts was particularly striking. It was noted that:

- (a) All those who achieved 'A' overall achieved only 'B' in the exam.

- (b) All those who achieved ‘B’ overall gained an ‘A’ mark in the Extended Response/Paper 2 and a ‘C’ in the exam.
- (c) All those who achieved ‘C’ overall also gained an ‘A’ mark in the Extended Response/Paper 2; however, they all failed the exam.
- (d) All but two of those who failed overall nonetheless gained an ‘A’ mark in the Extended Response/Paper 2. (The remaining two scored a ‘B’ mark.)

The sample is, of course, far too limited to justify seeing this pattern as a general phenomenon. In any event, the panel took the view that the notably better performance in the Extended Response/Paper 2 was in line with the intention that this should give candidates the opportunity to show their best work.

Exam (Paper 1)

The panel felt that the same grades show comparable performance throughout. The following points may additionally be noted:

Candidates who scored lower marks tended to do so either because they (apparently) ran out of time or because they were limited by the amount of recall they could deploy in their answers. (For further comment, see 5.2.)

In one case a candidate achieved a C overall (19 for the exam) despite failing to attempt a context; in another case a C was achieved (with 19 for the exam) despite failing to attempt two contexts. Managing time was an issue—for example, for candidates who tried to deploy material from their Extended Response/Paper 2 in the exam.

The way the variety of types of ‘source comparison’ question were answered was examined. The sample scripts did not show any clear indication that questions of a certain type were easier or more difficult than others, or that this had any impact on comparability of performance across the years under review (for example, if candidates had chosen contexts where one type of source comparison question was predominant rather than other types of question). For further comment, see 5.2.

5.2 Are there specific weaknesses and strengths in the 2003 and 2006 scripts compared to 2007 scripts at the same level?

The panel noted a pattern of performance experienced by candidates across all years. In sum:

- (i) The highest mark was for the Extended Response/Paper 2 (see 5.1).
- (ii) The most demanding skill was recall. The ‘describe’ question in each context tended to be answered more poorly than others, especially by the weaker candidates, who also tended to struggle with recall in the source analysis question.
- (iii) There were no examples in the sample of a candidate who performed well in all three contexts. Performances tended to tail off in the last context, or only the first context was tackled convincingly.
- (iv) In the source comparison questions of the type ‘How far do they agree/disagree’, only the better candidates were able to differentiate this from

the 'Compare' type questions. The weaker candidates tended to discuss provenance (which is irrelevant).

- (v) In source comparison questions generally it was also noted that there were many instances where candidates quoted sources as well as commenting on them: only comments are required. In source comparison questions generally it was also noted that there were many instances where candidates quoted entire sentences from sources as well as commenting extensively on them when they could have saved considerable time by giving the judgement and paraphrasing the relevant part of the sentence.

The only aspect where a difference was detected between 2007 and previous years is that, in source comparison questions of the type 'Compare as evidence' or 'Compare the views', the provenance of sources was included in answers in 2007 (and also in some answers to questions of the type 'How far', when this is not required), but was not dealt with so effectively in 2003.

Final statement: explaining the increased pass rate in 2007

The panel noted that the pass rate had increased in 2007, and observed that this could not readily be explained simply by the marked change in the nature of the population of candidates (see Introduction). At the same time, the panel found no evidence that standards had changed to any significant degree in the curriculum, the exam questions, the assessment instruments and arrangement, or in marking criteria. How, then, can the increased pass rate in 2007 be explained? A crucial additional piece of information is that the average mark in the Extended Response/Paper 2 (marked out of 20) had risen from 11.2-11.7 in 2004-6 to 13.3 in 2007; the increase from 2006 to 2007 was 1.8. By comparison, the average mark in the exam (marked out of 50) had fluctuated in the same period from 25.9 in 2004, to 28.2 in 2005, to 26.4 in 2006, to 28.1 in 2007. These figures point clearly to 2007 as a year of unprecedented achievement in the Extended Essay, at least in relation to 2004-6, if not earlier. This would partly account for the higher pass rate. It will be recalled (see 5.1) that the Extended Essay consistently rescues candidates who have failed the exam. However it is encouraging that the average mark for the 2007 exam has risen by 1.7 from 2006. We must await the result of the 2008 and subsequent examinations to discover if this is part of a trend (perhaps due in part to the greater confidence and experience of teachers, or to continuing in-service) or if 2007 has been a particularly able cohort.

The question of whether the increased pass rate in 2007 signals a decline in standards or not can therefore be ascribed mainly to the Extended Response. Has this got easier? How else can the increase in the average mark for this element be explained? The panel saw no evidence that candidates found this easier. What the panel found was that the marking instructions had changed in such a way that it had now become easier for existing criteria to be applied consistently (see 4a). The panel felt that this was likely to make markers more willing to award marks in the 'A' range, particularly at the top of the range. The removal of awareness of different possible interpretations as a criterion for an 'A' mark could, theoretically, also have contributed to this. The firm conclusion of the panel was

that the new marking instructions do not represent a change in standards, but signify no more than an important change in procedure which is designed to ensure that criteria are applied properly, and that unintended marking practice is minimised. If the average mark in the Extended Response, and in the overall pass rate, has increased as a result, then this should be welcomed.