Dear Sharon, Gráinne and James,

**A Level Ancient History (H407)**

Thank you for your letter sent on 19 September. We have sent you a separate response to the second part of your letter in which you request information about examiner recruitment for A Level Ancient History under the provisions of the Freedom of Information Act (2000).

Your letter covers a number of different processes, and I thought it would be useful to provide some information to help shed some light on them before directly answering the questions you have posed.

Firstly, given your letter covers both marking and awarding I wanted to be clear as to what we mean when talking about the two areas as I often see the two words being used interchangeably.

- **Marking**: in crude terms, a process that effectively turns candidate answers into a numerical value thereby creating a rank order, at question level and for the whole question paper, based on how well the answers meet the assessment criteria.

- **Awarding**: also referred to as *grading*: once the rank order has been established, the awarding process identifies appropriate points within that rank order to demark differing levels of performance – i.e. placement of grade boundaries.

Given that marking is the first part of the process, I thought I would address it first. It should also help contextualise any information we release under the Freedom of Information Act.

**Marking**

At the start of each standardisation meeting, there is an admin briefing and it outlines the key requirements of the purpose of standardisation:

> “… all examiners must have a well-founded and common understanding of the requirements of the mark scheme and can apply them reliably.”

We judge the experience of our examiners on a much more holistic criterion than just whether they have taught this version of the specification or a particular depth study. It is essential that our examiners can apply the mark scheme and marking grids consistently and reliably and have detailed subject knowledge so that we can establish an accurate rank order. As well as standardising our examiners, we test the accuracy of their marking throughout the marking period. We have an experienced team of examiners who do mark consistently and reliably, and many have done so for the decade or so since I have been involved with the A Level Ancient History qualification. Post-series reviews have not raised any concerns about the quality of marking.
How do we ensure that a comparable standard is set across the different papers?

There are several stages to ensuring that there is comparability between the different A Level Ancient History papers.

Through the question writing process there is currently a common test constructor (who revises the question papers) and a common test coordinator (effectively an oversight role). This is our first mechanism to ensure that each of the six papers has a comparable level of difficulty.

The second stage is through our approach to marking. We have one Principal Examiner for the three Greek papers and one Principal Examiner for the three Roman papers. This is so they can ensure that a common standard is set across the three different papers. Furthermore, we have built-in cross-standardisation to ensure that there is a common standard across both the Greek papers and the Roman papers. At the standardisation meetings for the qualifications, both Principal Examiners are present. They not only understand the standard set but also, to date, have marked to that standard when they have been an assistant examiner for papers for which they are not the Principal Examiner.

The final stage, and one that we haven’t needed to use, is through the awarding process. As the different depth studies are assessed in different question papers, we can award different grade boundaries for the different options if there is strong evidence to suggest that one or more papers haven’t performed in line with others.

The awarding process

As detailed above, the awarding of grade boundaries is a separate process from that of marking and takes place at the end of the marking period.

For all qualifications, we set the grade boundaries at ‘key judgemental’ grades at specification level. For A Level, these key judgemental grades are A and E. These are boundaries where the performance standard is known. Once the A and E grade boundaries have been set, the grade boundaries for grades B, C and D are calculated arithmetically so that the number of marks between them is even. The grade A* is then calculated as per the 2023 data exchange procedures which are detailed on the Ofqual website. This means that the A* grade boundary is also a key grade boundary that is selected and one which we have oversight of but does not require examiner judgements.

For reference, the key judgemental boundaries for GCSE are grades 7, 4 and 1. GCSE grades 6, 4, 3 and 2 are calculated arithmetically. Just like A* in A Level, the grade 9 in GCSE is calculated according to the Ofqual data exchange procedures, and the grade 8 is then calculated arithmetically as a midpoint between the grade 9 and 7 boundaries.

When awarding the grade boundaries, a range of statistical information is considered as well as examiner judgements. The statistical information that is considered includes:

- Prior attainment information of the candidates who have taken the qualification. For A Level, the measure of prior attainment is the profile of mean GCSE scores.
- A prediction of the expected outcomes based on outcomes in previous years and the ability of the cohort (as measured by their prior attainment) to ensure comparability across time and between exam boards (where appropriate)
- Mean marks and standard deviations for the current examination series as well as previous examination series
- Unit pair analyses are also conducted where relevant to the qualification structure. This looks at how common candidates performed, at cohort level, in two different papers.
Examiner involvement

The awarding process involves statistical data as well as involvement from senior examiners. As explained in this blog from Ofqual, when carrying forward standards in the first year of the new qualification, “Exam boards will be relying heavily on the statistical evidence to do this, but also using senior examiners to check the grade boundaries that the statistics are pointing to.” This was also true this year when realigning to the pre-pandemic standard. Examiners’ inputs are also helpful when determining appropriate notional grade boundaries for each paper and also at the E/U boundary where the statistical evidence is usually more open to interpretation.

Examiners familiarise themselves with archive scripts which were on the boundary marks for ‘key judgemental grades’ (grades A and E) in the previous year – for 2023, examiners looked at archive scripts from the 2019 series. By doing this they gain a sense of any difference in the demand of the question paper for the current year and the benchmark year.

Each member of the awarding committee will then review different scripts, where available, on and around the requested mark points to make a judgement about which mark points reflect each grade’s performance, and which falls below that standard.

These examiners’ judgements are just one of the sets of evidence that is used by OCR to aid them when setting the grade boundaries.

The comparable outcomes approach

All English exam boards use the Comparable Outcomes Approach when awarding GCSE, AS and A Level qualifications.

The aim of comparable outcomes, as detailed in this Ofqual blog, is to ensure “that if the national cohort for a subject is similar (in terms of past performance) to last year, then results should also be similar at a national level in that subject.”

The comparable outcomes approach is cohort-referenced (outcomes based on a fixed relationship with the statistical profile of the cohorts). It is not norm-referenced (outcomes based on a fixed distribution) or criterion-referenced (use of performance descriptors or other fixed criteria for comparison).

If, for example, for whatever reason, marking was harsher in comparison to previous years, and all other factors remained constant the comparable outcomes approach would ensure that the same outcomes are achieved, by setting lower grade boundaries.

How does comparable outcomes work?

Exam boards create a prediction of the expected outcomes for a cohort based on that cohort’s prior attainment and comparing it with the prior attainment and outcomes for a reference year for that subject. If the prior attainment for a given year was exactly the same as the reference year, one would expect outcomes at cohort level, to be very similar. However, if the profile of the cohort was different to the reference year, this would be taken into account. For example, if the cohort’s prior attainment was greater than the reference year, this would be reflected in the predictions, and more of the top grades would be expected.

For qualifications where all of the papers are mandatory, the aim would be to find the overall specification level marks for the matched candidates which align closest to the cumulative percentages suggested by the prediction matrix for grades A and E, and then work through the procedures outline in the Data Exchange document for the grade A* boundary. However, for those qualifications like A Level Ancient History which have optional papers and optional pathways, we would first need to analyse the relative performance of students across the different papers.

The A Level Ancient History qualification has been designed so that one can analyse candidates’ overall performance in Section A of the papers which is common across the three
optional Greek papers and the three optional Roman papers. This can also be done through the analysis of candidate performance on the mandatory interpretation question. Where the analysis points to candidates performing in a similar way across papers, there is a rationale that common grade boundaries should be set for the three papers.

If there is evidence to support common boundaries across all three Greek and all three Roman papers, we would find the overall specification level marks for the matched candidates which align closest to the cumulative percentages suggested by the prediction matrix for grades A and E. From there, the grade boundaries for grades B, C and D would be calculated arithmetically as detailed earlier on within this letter. We would then work through the 2023 data exchange procedures to calculate the A* grade boundary. For A Level Ancient History in summer 2023, the upshot of working through the data exchange procedures was that we had to find the overall specification level marks for the matched candidates which aligned closest to the cumulative percentages suggested by the prediction matrix for A*.

After grade boundaries are selected at specification level, we go about creating grade boundaries for each paper. This helps give teachers and students an idea of how they performed at paper level. These are the grade boundaries that are available on the OCR website, and have been replicated below for reference:

<table>
<thead>
<tr>
<th>A Level Ancient History</th>
<th>Max Mark</th>
<th>a*</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>H407 11 Sparta and the Greek world</td>
<td>98</td>
<td>82</td>
<td>71</td>
<td>57</td>
<td>44</td>
<td>31</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>H407 12 Athens and the Greek world</td>
<td>98</td>
<td>82</td>
<td>71</td>
<td>57</td>
<td>44</td>
<td>31</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>H407 13 Macedon and the Greek world</td>
<td>98</td>
<td>82</td>
<td>71</td>
<td>57</td>
<td>44</td>
<td>31</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>H407 21 Republic and Empire</td>
<td>98</td>
<td>83</td>
<td>72</td>
<td>59</td>
<td>45</td>
<td>31</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>H407 22 The Eleven Caesars</td>
<td>98</td>
<td>83</td>
<td>72</td>
<td>59</td>
<td>45</td>
<td>31</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>H407 23 Emperors and Empire</td>
<td>98</td>
<td>83</td>
<td>72</td>
<td>59</td>
<td>45</td>
<td>31</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>All options</td>
<td>Overall</td>
<td>196</td>
<td>165</td>
<td>143</td>
<td>116</td>
<td>89</td>
<td>62</td>
<td>36</td>
</tr>
</tbody>
</table>

As you can see, the awarding of grade boundaries is a complex task but I hope it gives you some useful information in order to understand the process and help contextualise my answers to the direct questions you have posed.

**Why are outcomes lower in Ancient History compared to OCR History and OCR Classical Civilisation?**

The outcomes for OCR’s Ancient History, Classical Civilisation and History qualifications were each calculated using the subject-specific predictions based on the relationship between prior attainment at GCSE and performance in A Level for that subject’s cohort, in each subject.

One of the explanations for the differing outcomes at cohort level relates to the profile of cohort’s prior attainment.

The cohort sitting A Level Ancient History had significantly lower prior attainment than the cohort sitting OCR’s History qualification (H505) and the cohort sitting OCR’s Classical Civilisation (H408) qualification, and therefore shows the A Level Ancient History cohort, in terms of prior attainment, is generally speaking not as strong. Additionally, the Ancient History cohort has got weaker over time, and this shift has lowered the predicted number of A*s in recent years.

We are confident that the awarding process for A Level Ancient History was accurate and fair and was consistent with Ofqual’s ‘comparable outcome’ approach. We are also confident that boundaries were set in the fairest way possible across all components to achieve outcomes which were in line both with predictions and with previous series.

However, we are aware of the concerns that have been raised by teachers regarding the equality of outcomes between History and other subjects and Ancient History which has been mentioned in your letter. These concerns have been raised with our Research and Technical Standards team and they will conduct further analysis of the available data.
The ‘B cumulative’ percentage for Ancient History is very similar to Classical Civilisation (59.16 / 60.59), and yet there is a 7.78% gap between the ‘A cumulative’ percentage (19.40 / 27.18), and a 2.85% gap between the rates of A* (2.75 / 5.6). How do you explain this?

The ‘B cumulative’ percentage for Ancient History is 2.43% greater than for History (59.16 / 56.73), and yet there is a 5.5% gap in History’s favour between the ‘A cumulative’ percentage (19.40 / 24.9), and a 2.85% gap between the rates of A* (2.75 / 5.6). How do you explain this?

As detailed above, the key judgemental grade boundaries which are selected are the grade boundaries for A and E. The grade boundaries for grade B, C and D are calculated arithmetically. The grade A* boundaries are selected according to Ofqual data exchange procedures. Looking at the grade boundaries for OCR’s History and OCR’s Classical Civilisation qualifications, the grade A* boundaries for both qualifications were selected to ensure matched candidate outcomes were closest to the predictions.

The differences in Grade A* and Grade A cumulative percentages result from the more able Classical Civilisation and History cohorts as described above. Given that the grade B boundary is calculated arithmetically, the much smaller difference between the two qualifications is simply a function of the distribution of marks.

Why has Classical Civilisation and History A* and A cumulative percentages increased notably since 2019 while Ancient History has increased fractionally on its A* percentage from 2019, and decreased on its ‘A cumulative’ percentage?

The changes identified are linked to the differences in the cohort profiles for the qualifications in question in 2019 and 2023 as described earlier. Given that the question refers to Grades A* and A, the outcomes would have been the result of being as close to the predictions as the relevant A Level Ancient History, A Level Classical Civilisation and A Level History prediction matrices suggested.

With reference to the point above, how is it that the Ancient History A* percentage has increased slightly, but ‘A cumulative’ percentage has decreased? You are saying that there will be more candidates who will reach A* standard than in 2019, but fewer who will reach A standard. Is this realistic?

Yes. The different cohort profiles from 2019 to 2023 meant that this is the case.

It is our understanding that no A*s were awarded to candidates that entered both H407/13 (Macedon) and H407/21 (Republic), which is entry Option G. Is this a realistic outcome? Are you satisfied that candidates offering this Option are being examined as consistently as candidates offering other Options?

Yes. As described above, we have a comprehensive system in place to ensure that standards are aligned across all optional papers, so I have no doubt that the candidates were examined consistently. All examiners mark scripts anonymously and the examiners marking H407/13 and H407/21 do so not knowing which entry option the candidate has entered.

For those students taking Option G, there were some candidates that received a notional a* in the Macedon paper and the Republic paper. However, the cohort taking Option G generally did not perform consistently well on both papers, whereas candidates entering other options did. With the A* grade boundary set at 165 marks, a slightly weaker performance on one of the two papers means that they would have had to do extremely well on one to get an A*.
Many centres are reporting that there was a significant difference between the marks their candidates scored on the Greek paper and the Roman paper, with the Roman paper often being 10 or more marks higher. How great is the difference between the average (both mean and median) for each paper?

Please find below the information you have requested, which shows that the mean mark for the Greek papers was higher than the mean mark for the Roman papers.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>H407/11–13</td>
<td>60.06</td>
<td>62</td>
</tr>
<tr>
<td>H407/21–23</td>
<td>57.49</td>
<td>60</td>
</tr>
</tbody>
</table>

Thank you for detailing the concerns of teachers, and also sharing the survey which was used to help draft the letter. The awarding process is complex, and intersubject comparability is even more complicated. Therefore we would be happy to set up a meeting with you so that a colleague or colleagues from our Research and Technical Standards team can talk to you about this matter. I’d also be happy to widen the invite to include the Classical Civilisation, Latin and Classical Greek subject representatives. Please let me know if you’d like to take us up on this offer.

I hope that this letter goes some way to reassure you.

Yours sincerely

Alex Orgee

Subject Advisor – Classics