**OCR B (SHP): A case study of summative assessment.**

**Introduction**

Like many other History departments, in our school we are required to set Year 10 a summer exam, mark it and award each student a summative grade. This can be tricky at the best of times, what with lack of time for in-depth standardisation and end-of-year exhaustion (both the students and us!). This year, we of course had the added challenge of a new course, new exams and a new grading system to wrestle with. As such, we decided to go all-out and, working with OCR, replicate the real-life standardisation process as closely as possible. In this blog I will attempt to give you an overview of the process which we followed, plus – perhaps more importantly – some thoughts on how the experience and lessons learned might shape our future teaching of the SHP B course.

**Context**

Our school is a more or less typical, large comprehensive school. We currently have 5 classes in Y11 (last year’s Y10) – a total of 125 students. We follow the OCR B SHP course and have opted for units on The People’s Health, The Norman Conquest, The Making of America and Living under Nazi Rule; plus a History Around Us study on Conisbrough Castle. We decided to set the OCR Practice Paper 1 for The People’s Health and the Norman Conquest – available on Interchange – as the Y10 exam.

As for myself, I am a History teacher within our department but I also an OCR Principal Examiner and exam author, which made it a bit easier for me to help steer the department through the process of what happens once an exam has been taken by a cohort of students. I exchanged ideas with other examiner colleagues throughout the process which is outlined below.

**The process**

1. The students sat the exam and we collated 3 responses from each teaching group – in theory what should have been a top, middle and bottom response.
2. The published generic mark schemes are great in terms of providing a hierarchy of types of responses. The trouble is that when we have descriptors which set out requirements like ‘sophisticated’ or ‘sound’, they are open to interpretation. We needed to make it clearer exactly what kind of response we thought would qualify as ‘sophisticated’, ‘sound’, etc. So using the published generic marks schemes, plus these 15 responses, I wrote mark schemes for each question which the students would answer and got some feedback from OCR examiner colleagues. This gave us a practical set of schemes which were tailored to the specific questions, interpretations and took into account the cohort’s overall performance.
3. I then selected 5 sample scripts across the range of performance, discussed them in detail with other OCR colleagues, and definitively marked them according to the mark schemes I had devised, which were then tweaked as I came across things that didn’t work. I’m not going to lie – this process took about 10 hours to complete and so does involve a heavy time commitment if it’s to be done properly! This was one reason why we thought writing up this exercise was a good idea – to hopefully save some other teachers the same 10 hours worth of pain!
4. The mark schemes and 5 sample scripts were then distributed to the other Y10 teachers in the department who marked them (separately).
5. We then all came together and ran a mini standardisation meeting: for each question, we read through the mark scheme and shared the reasoning behind it. Then the teachers shared the marks they had given for each script and further discussion ensued. In some cases we revised the final mark awarded for a particular response, as would happen at a real standardisation meeting. Again – the time it takes to go through this cannot be underestimated. Our SLT allowed us to have 3 hours off timetable for our meeting, and even then, we did not finish reading all the responses together! However, I would argue that the insight and experienced gained through this process was invaluable.
6. The Y10 teachers all went away and marked the rest of their own class’ responses.

**Some caveats**

Our thoughts and findings from carrying out this activity are outlined below and hopefully you will find them interesting and useful. Before you read them, however, it is vital that the following caveats are noted:

1. The mark schemes that we have written cannot and should not be viewed as the definitive mark scheme for this or any other exam. The writing and implementation of any mark scheme is dependent on a) the particular question being asked and the sources/interpretations which are being used and b) the performance of the cohort (see points 2 and 3 below). Most importantly, the final mark scheme is the decision of the relevant Principal Examiner – although of course we hope that they will use similar thinking.
2. Like most schools we are required to give our students grades, despite the fact that an entry of 125 in one school does not really give a large enough sample and of course there are no past cohorts to make comparisons with. At the time of writing we are yet to make a decision as to how to do this.
3. I don’t know if this is the case with your Y10s, but we generally find that ours do not revise in the same way for their mocks that they will for the live exams. Many do not revise at all, and those who do are often overly-reliant on the CGP revision guides with which they are issued. Added to this, we did no in-class revision of actual content with them. All of this results in the mean mark and overall performance being much lower that it will (hopefully!) be by the end of Y11. So again, this had impacted on the mark schemes produced and the grades which will be awarded.
4. As a department we feel that (granted a 9-day week and 28-hour days … ), the process would still have benefited from some post-marking moderation to ensure our data was even more consistent.

**Summary of responses and findings**

You can find attached a copy of Practice Paper 1 plus the mark schemes which we used.

**The People’s Health**

*Question 1*

Quite often, our students did not answer along the lines suggested in the indicative content. For example, in 1a (‘Name one threat to people's health in medieval towns’), the vast majority of ours wrote ‘The Black Death’. Or, for 1b (‘Give one example of a law passed to improve public health in the nineteenth century’), our candidates did not name specific Acts but rather described the effects of such measures, eg ‘they built sewers’ or ‘they cleared up waste’. Given that these questions are intended to be very gentle, introductory questions, we took a very lenient line on what was allowable.

*Question 2*

Here we allowed for basically two types of response. At Level 1, responses basically listed valid local or national government responses to plague with no successful attempt to organise their response around a second-order concept. At Levels 2 and 3, candidates organised their response around a valid second order concept. Typically (and despite the local/national possibility offered by the question itself), this was done through causation (ie reasons for particular actions being carried out) although a few candidates were successful in describing how response changed and/or stayed the same across the period. In general we found that a lot of these answers were valid but very vague and lacked precision and specific examples.

*Question 3*

The wealth of the monasteries as an underpinning factor did not come up at all in responses. The majority listed factors such as water supply, sanitation etc. These were marked at Level 3. Better answers explained how/why those factors led to good health and achieved Level 4 (one reasons explained) and L5 (two reasons explained). There were quite a lot of irrelevant responses along the lines of monasteries having infirmaries. These were marked at Level 2 or below.

*Question 4*

Many students found it hard to decide whether it was a question about attitudes to health, or a question about the impact of measures which were taken. However – as was the case with both essay questions – by far the biggest barrier to students was not having precise knowledge and specific examples at their fingertips to deploy. The responses were overwhelmingly vague and woolly and as such the many responses were stuck at Level 1 or Level 2. Better responses were able to explain the attempts of 1 period to improve public health by pointing to one or more examples (Levels 3 and 4) and the best ones compared the 2 periods (Levels 5 and 6). The top mark was awarded within Levels 5 and 6 if the student provided a valid clinching argument.

*Question 5*

This question was challenging because of the need to compare c.1900 living conditions to later conditions in order to answer the question properly. The nature of the question (which was centred around ‘changes’ in living conditions as opposed to ‘improvements’) meant that it was only fair to allow completely one-sided arguments access to the top levels, given the difficulty of arguing for lack of change since 1900. However, other than that, the mark scheme followed a similar structure to that of Question 4. All the comments above about vagueness and lack of precision were also true here, although more students were able to access a low Level 3 mark because they could fall back on general descriptions of today’s living conditions.

**The Norman Conquest**

*Question 6a*

There was no real problem for students identifying one way in which the illustrator tried to make women look important but they found it very difficult to develop this. In the end we allowed marks for identifying several ways as an alternative approach.

*Question 6b*

This proved tricky for the students. The vast majority of our students tended to see the interpretation as a real scene here and said things such as, ‘I would investigate what kind of jobs this woman did’. As such they ended up in AL1 because they had not identified a valid line of enquiry. A handful did use Q6a as a cue and reached Levels 2 or 3 by identifying an enquiry based around the role of women. However, very few of our students truly got to grips with this question and there is work for us to be done here in terms of helping our students understand the nature of valid historical enquiry.

*Question 7*

On the whole our cohort did relatively well on this question. There were not many who were stuck at Level 1 (assertions or simplistic comparison of provenance) or Level 2 (comparison of isolated details). The majority of them got to the idea of how Hereward was portrayed and so did quite well, achieving Level 3. They struggled a bit more with explaining the differences. However, on the whole they got the purpose of Interpretation B and therefore why Hereward was represented so dramatically. A pleasing number reached Level 4. With Interpretation C, very few got beyond pretty feeble comments along the lines of it being aimed at children. One unforeseen difficulty was that some students dug further into Interpretation C and argued quite successfully that it was not really very different from B – Hereward emerges as a hero who can only be defeated by betrayal. This left them with struggling to explain why they differed.

*Questions 8 and 9*

If possible, even more so that with the People’s Health section, our students’ lack of revision showed itself on these essay questions. Again, answers lacked specific knowledge and frequently lapsed into assertion and vague (and often inaccurate) description. With Question 8, particularly, students struggled to make explicit comparisons with Anglo-Saxon England. Q9 was the more popular of the two questions but with both questions, a huge portion of responses failed to get past Level 2.

**Conclusions**

1. *The role of knowledge selection*: It was clear from many of our students’ responses, especially in the essay questions, that they simply did not have enough knowledge to be able to answer the questions with the relevant examples and explanation necessary. However, that is not to say that students who have lots of knowledge will automatically be able to deploy it effectively. Indeed, GCSE candidates often make the mistake of throwing all they know at each question and hoping that some of it sticks. The essay responses themselves do not need to be full to the brim with lots of ‘facts’. Indeed the handful of excellent responses which we did get had only one or two well-chosen and relevant examples in each section. The problem is that, to be able to have just one or two relevant examples, you need to be able to select them from a wider array. As Gustave Flaubert said, ‘writing history is like drinking an ocean and peeing a cupful.’ I think that our revision schedule next year must necessarily include not just factual recall tests, but activities which allow students to practise this skill, eg (‘Read the answer … can you tell what the question was?’ … ‘Here’s a list of facts; they are all accurate but which ones are relevant to Question X?’).
2. *Second order concepts*: Question 6b seemed really alien to our students which maybe points towards their lack of understanding of the nature of historical enquiry. Some further practice at generating valid questions throughout the whole course is much needed here. It may also be worth pointing out to our weaker students that if they are really stumped, Q6a will usually assist them in directing their thinking.
3. *Revision of change in the thematic unit* – I think we may have not emphasised enough what was special or unique about each time period in the People’s Health and placed too much emphasis on continuity. I say this because it was often apparent from students’ responses that they had a very confused understanding of, for example, responses to the Black Death as opposed to the Plague; or measures passed in the 19th C as opposed to the Early Modern period.
4. *Interpretations* – It was pleasing to see that the majority of our students were able to examine things like the author’s purpose and its impact when writing about why interpretations differed. However, a minority still stuck to undeveloped comments relating to provenance (eg ‘one is a history book for children so it has a young audience’) so I think I’ll be searching round for some more meaty interpretations with a clear identifiable feature for them to practise on. Commemorations or personal/geographical connections between the author and the events portrayed usually work well.