Romeo and Juliet – technical explanation

24 August 2017

Our GCSE English Literature paper (J352/02) on Poetry and Shakespeare had an error on it. This error was in one of two Romeo and Juliet questions. The question asked about Tybalt's hatred of the Capulets when it should have asked about his hatred of the Montagues. We are very sorry this error was not picked up by our quality assurance procedures and apologise for the distress it has caused students, parents and teachers.

In dealing with this error, we wanted to minimise any disadvantage to students as much as we possibly could. This document describes the approach we used to achieve this and some of the key issues we considered.

How did the error happen and why was it not spotted before the exam?

Each of our exam papers goes through a lot of checks. This paper was checked by subject and assessment experts. Part of this process included an exam simulation. When reviewing the exam, the checkers saw the question as it should have been written rather than as it appeared on the paper. We will be implementing new procedures to prevent this happening in the future.

How have you dealt with this error?

1. The lead marker looked at a large sample of student answers.
2. The lead marker identified different response types to the question with the error:
   - Some students had not been affected at all and had answered the question as it had been intended.
   - Some students had answered the question as it had been written, addressing Tybalt's hatred in general.
   - Some students had started to write an answer to this question and then changed to the other Romeo and Juliet question.
   - Some students had answered the other Romeo and Juliet question instead.
3. The lead marker amended the mark scheme so students would get marks if they had written about Tybalt's hatred in general or if they had written about his hatred towards either family. As long as students wrote about his hatred and fulfilled the other requirements of the mark scheme, they were credited with all the marks they would have received if they had written about Tybalt's hatred of the Montagues.
4. All the examiners for this paper were then standardised according to this approach by the lead marker. After standardisation, they had to pass our usual standardisation tests before they were allowed to mark.
5. During the marking of this question paper, we carried out our normal, extensive checks to make sure all examiners were marking correctly and to the right standard. We also asked team leaders and the lead marker to check all answers to the Romeo and Juliet question to make sure all examiners had fully understood the amended mark scheme.
6. We used wide ranging statistical analyses to determine how students who answered the Romeo and Juliet questions did in comparison with students who had done the other unaffected

Shakespeare questions (on Macbeth, The Merchant of Venice, and Much Ado about Nothing). This told us:

- Overall, the question paper with the error worked as a valid and reliable assessment. However, the amended mark scheme had only reduced rather than removed the impact of the error. On average, students who had answered the question with the error had overall performed slightly less well (by approximately two marks) on the whole question paper.
- On average, students who had answered the other Romeo and Juliet question had also done slightly less well on the whole question paper, but only by one mark.
- Some students had not been impacted at all, achieving full marks on the Romeo and Juliet questions, and the paper as a whole.
- What we did not know, and cannot know, is the exact impact (in marks) of the error on each *individual* student.

After evaluating the impact of the error, we determined an approach to mitigate its impact as much as possible for individual students:

1. We 'assessed' marks for the Romeo and Juliet questions for all students who answered either question. This is a tried and tested method which all exam boards use when calculating (=assessing) marks for students who have permitted partial absences or special considerations.

2. In order to assess the marks for students answering the Romeo and Juliet questions, we first analysed the performance of students who answered a question on the other Shakespeare plays (Macbeth, The Merchant of Venice and Much Ado about Nothing). This allowed us to determine the correlation of their marks on paper 1 and the poetry section of paper 2 with their mark on the Shakespeare question. Our statistical analysis demonstrated a strong correlation between these students' marks in paper 1 and the poetry section of paper 2, and their marks on the Shakespeare question.

3. For students who answered questions on Romeo and Juliet, we then used the correlation to calculate the 'assessed' mark for the Romeo and Juliet question from their marks on paper 1 and from the poetry section of paper 2.

4. We have given each student whichever mark is higher – either 'assessed' (statistically calculated by us) or 'examined' (given by the examiner). Overall, we believe this is the fairest method of dealing with the error.

Why didn't you simply add two marks for every student who had done the Romeo and Juliet question with the error and one mark for every student who had done the other Romeo and Juliet question?

We knew that would be unfair because some students had full marks for those questions anyway, so it would not have made sense to add more. It would also be unfair to those students who had been affected much more by the error and lost more marks than one or two.

How can you compare the Romeo and Juliet question(s) with the other Shakespeare questions in the paper?

The question paper J352/02 has two sections: section A, which is poetry; section B, which is Shakespeare.

Section B contains a choice of two questions on each of the following plays; Romeo and Juliet, Macbeth, The Merchant of Venice, and Much Ado about Nothing. For each play, the first question is an 'extract' question, where students are asked a question based on a given extract from the play. The second question is a discursive question only, without an extract (a non-extract question).

The paper is structured in this way so the questions for students are comparable, regardless of which Shakespeare plays they studied. This means we expect students of similar abilities to perform very similarly indeed on all the extract questions, and likewise on all the non-extract questions. After the examination, we can compare students' performance for the extract and non-extract questions on the different plays.

The error in the Romeo and Juliet question was in the actual 'extract' question (question 4) – but in the question part, not in the extract itself.

How many students answered questions on Shakespeare?

Total number of candidates who answered Shakespeare questions: 14,250

- Macbeth answered by: 6,605
- Romeo and Juliet answered by: 5,651
- The Merchant of Venice answered by: 1,265
- Much Ado about Nothing answered by: 466

263 students made no visible attempt to answer any Shakespeare question.

How many students answered each of the Romeo and Juliet questions?

Of the 5,651 students answering a question on Romeo and Juliet, 4,299 answered the question with an error (Q4, the extract question), and 1,352 answered the unaffected question (Q5, the non-extract question).

How many students answered the question with an error (Q4, the extract question) as it was intended?

About two thirds of students answered question 4 as it was intended, in other words writing about Tybalt's hatred of the Montagues.

Were there a lot of students who started to answer question 4 and then moved to question 5 instead?

There is no statistical evidence that the error in question 4 encouraged students to move to question 5 instead. The proportion of students answering non-extract questions is much greater with Macbeth and The Merchant of Venice.

Did some students decide to answer a question from a different play instead?

There was no evidence students moved to a question on a different play because of the error on question 4.

What if students started with the Romeo and Juliet question and answered the poetry questions last?

355 students answered either question 4 or question 5 before the poetry questions. This was almost the same proportion of students who answered other Shakespeare questions first and answered the poetry questions last. We found no evidence that students who started with the Romeo and Juliet section were more affected by the error than those students who started with the poetry section.

Did the error mean some students did not answer any Shakespeare question?

Our analysis suggests around 50 students were unable to answer any Shakespeare question as a result of the error. We have assessed marks for all students who did not write an answer on the Romeo and Juliet section who we identified as having been taught Romeo and Juliet at school.

How many students gained full, or nearly full, marks on the question with the error?

More than 75 students achieved full marks in the Romeo and Juliet questions and nearly 300 were awarded 37/40 marks or more, before any adjustments were made.

Why did you not just discount the question with the error or discount the whole of the Romeo and Juliet section?

We did consider this option. However, during the lead marker's initial review of scripts, it became clear there were a lot of students who had answered this question really well. We did not want to discount all of their hard work and the preparation they had put into revising for the Shakespeare section. From the statistical analysis, it is also clear question 4 worked validly and reliably as an assessment.

How have you set grade boundaries for GCSE English Literature?

We did **not** set different grade boundaries for different groups of students. We have set grade boundaries in the usual way and according to the Ofqual requirements for reformed GCSEs: we needed to get the judgemental grades (grades 7, 4 and 1) as close as possible to predictions for all students who have taken the qualification. We have selected grade boundaries that fulfil this requirement for all students, including those who took the Romeo and Juliet questions.

Are there any external measures we could use to compare with students' marks in this paper?

Yes. Key Stage 2 average scores and GCSE English Language correlate well with GCSE English Literature. We calculated correlations with both of them for students who answered the Romeo and Juliet questions and for all other students. On average, the relationships between either students' KS2 scores or their English Language results with their English Literature results are as we would expect them to be had the error not been there.

Our school's results look even more unexpected than we had anticipated from the new qualification – surely that's because of the error in the paper?

We investigated whether the error might make results more variable than expected for schools. As such, we analysed results for OCR centres who had taken our GCSE English Literature in 2016 and 2017. This showed the levels of variability are well within the range we typically see amongst established GCSEs. So any variability looks to be the result of the new qualification, rather than the error. Please also have a look at Ofqual's publication on variability of results: https://www.gov.uk/government/publications/year-on-year-variability-in-school-level-gcse-results, and/or at research done by Cambridge Assessment on the same topic: http://www.cambridgeassessment.org.uk/Images/372751-volatility-happens-understanding-variation-in-schools-gcse-results.pdf.

What can students do if they want to find out more?

Students will have access to free copies of their marked exam papers. If students would like to query their results further, they are also able to request a review of marking as normal through Interchange. There will be no charge for these reviews for students who answered Romeo and Juliet questions.

How is an assessed mark calculated and how reliable is it?

We use a z-score calculation (an industry standard approach) to calculate assessed grades for a variety of situations, e.g. for some special considerations, for partial absences, in some malpractice situations. This technique has been used for many years by all exam boards where marks need to be assessed, and research has shown it to be a robust methodology.