

Ofqual Board

Paper 30/17

Date:

12 July 2017

Title:

National Reference Test – annual report

Report by:

Tim Leslie, NRT Project Director

Dennis Opposs, Standards Chair

Cath Jadhav, Associate Director, Standards and Comparability

Responsible Directors:

Marc Baker, Chief Operating Officer

Michelle Meadows, Executive Director, Standards, Risk and Research

Paper for discussion

Open paper (with a closed paragraph)



Issue

1. In July 2016, the Board approved the introduction of the National Reference Test (NRT) and for testing to be conducted in schools for four years, with the first tests to be held in schools in 2017. We indicated to the Board that we would report annually on the NRT. This paper is the first annual report.

Recommendation

2. The Board is invited to consider the paper.

Background

3. In 2013, the Secretary of State asked Ofqual to take forward the introduction of a national sample test to help set and maintain standards in the new GCSEs. During 2014, Ofqual developed detailed requirements for the test and procured a supplier for the design, development and delivery of the test. In April 2015, Ofqual awarded the contract to the National Foundation for Educational Research (NFER).

OFFICIAL – SENSITIVE

4. Following a successful full-scale operational trial in March 2016 (the “Preliminary Reference Test”), the Board approved the introduction of the test. We held the first live test in schools in February and March this year. We plan to hold the test each year. From 2019 (at the earliest), we will review the outcome of the NRT in the context of other information that is available to decide whether there has been a change in the performance of the national cohorts in GCSE maths and English language. If we conclude that there has been a change, we will agree with the exam boards how they will take this into account when they award the GCSEs.

Analysis

5. NFER reported the outcomes of the 2017 NRT to Ofqual in May. We have reviewed these and also taken advice from the NRT Sub-Group of Ofqual’s Standards Advisory Group (SAG), which met on 22 June. Their advice is summarised in Annex A. The SAG met on 30 June and endorsed the Sub-Group’s advice.
6. A summary of this year’s test administration and results is provided in Annex B. Ofqual attended 15 schools to observe the test taking place. There was a short delay to the start of the test in some schools because the school’s seating plan had not been co-ordinated with NFER’s test materials, which are personalised to each student. NFER is addressing this through revised procedures, including the scripted introduction used before the test starts, and will ensure all its test administrators are fully trained in the revised arrangements for the test in 2018. Although schools are familiar with invigilating exams, we need the NRT to be controlled in each school by NFER’s test administrators to ensure test materials remain secure and delivery is consistent.
7. Ofqual and NFER asked schools for their feedback on the administration of the tests. The overall response was very positive. Schools would like to have feedback on how their students performed. We do not consider it is feasible to do this as, without access to the test items, feedback would be of little value. Also, no student takes the complete test so individual students’ results would not necessarily reflect how they might perform in the GCSE.
8. Overall, this year’s test performed well and the results provide a good basis on which to measure possible changes in GCSE performance over time. We will consider whether results from 2017 or 2018, or in combination, should be used to form the baseline measure of performance. We do not need to decide this until after the test data is available in 2018.
9. The tests need to provide results that are comparable over time. We will monitor carefully for factors that could have an impact on this:
 - a. Changes in the willingness of independent schools to take part in the NRT; independent schools are not subject to the

legislation and therefore we can only request their participation on a voluntary basis. In the 2017 test, only 7 independent schools agreed to take part although some of those in the sample who did not take part were entering students only for an international GCSE and therefore were not eligible for the NRT.

- b. Changes in the profile of students who are asked to take the test but do not turn up; we achieved a significantly improved response rate this year (88%) compared to the test in 2016 (78%). Our analysis using KS2 scores of the students in the sample indicated that less able students were more likely not to turn up to the test. We will also use GCSE results, after awarding, to monitor possible changes in the student non-response profile.
 - c. Changes in the motivation of students taking the NRT; we have introduced the student survey to provide evidence of possible changes in students' motivation. This evidence, together with other information such as possible changes in attitudes of school leaders and teachers to the NRT, will be monitored.
 - d. Drift in marking standards; NFER uses the same mark schemes and trains its markers in the same way each year. However, there is a risk that, over time, markers may become more or less generous in the marks given. We are developing with NFER ways to monitor marker drift, for example through analysis of marks given to seed items.
 - e. We expect to be able change a small proportion of test items each year without losing comparability of results. (We may need to replace an item, for example, because its security has been breached.) We will use data from tests in 2017 and 2018 to validate the psychometric approach that enables items to be replaced without loss of comparability.
10. Ofqual continues to develop with the exam boards an approach to use the information from the NRT in their awarding of GCSEs. The current preferred option is to use NRT data to adjust the whole-cohort predictions for the judgemental grade boundaries, which the exam boards prepare from KS2 data. We expect to agree the approach with the exam boards by the end of this year and then hold a 'dry run' of the process in summer 2018. The Board has already confirmed that the NRT information should not be used in awarding in summer 2018. It will be in 2019, at the earliest, before it is used as evidence in awarding GCSEs. The approach to using the NRT outcomes will take into account other contextual information before concluding whether any change in cohort performance at GCSE is applied when GCSEs are awarded. Annex C provides an outline of the governance process Ofqual will follow to reach its decision. This was provided to the Board in May 2016.

OFFICIAL – SENSITIVE

11. We are confident at present that NRT outcomes can properly be applied to GCSE English language and mathematics, the two subjects assessed in the NRT. It is less clear how those outcomes might be applied to other GCSE subjects. They would not provide the same sort of direct evidence and yet may be indicators of changes of a construct sometimes called “general academic application” for the same 16 year olds. Exam boards will need to be mindful of the NRT results when awarding other GCSE subjects. Exactly how we will operationalise this will require further consideration.
12. We have started to prepare for the NRT in 2018. The tests will be held in schools between Monday 19th February and Friday 2nd March 2018. Our aim is to avoid making any changes in arrangements from those used in the 2017 test. The live items in the tests will be the same as those used this year. We plan to use the same size of school and student sample. Ofqual and NFER will write to schools that have been selected for the main sample on 18th September. We aim to avoid asking any school that took part in the 2017 test also to take part in 2018. We will continue to invite independent schools to take part if they are entering students for the GCSEs in either maths or English language. The statutory guidance to schools, subject to approval by the Secretary of State, will be largely the same as that issued for the NRT in 2017.

Finance and Resource

13. DfE provides additional programme funds to Ofqual to support the delivery of the NRT. In FY 2016/17, the project delivered within budget and DfE programme funds fully covered the costs incurred. In FY2017/18, we also forecast the project to be delivered within budget and for DfE programme funds to be sufficient.
14. The budget for FY 2017/18 is:

Pay	£255,000
NFER contract	£1,649,000
Other costs	£94,000
Total	£1,998,000

15. This section has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

16. This section has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

This section has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

Impact Assessments

Equality Analysis

17. We specified the NRT to offer the same access arrangements that students taking the NRT would be entitled to use for GCSEs. NFER provides these, in collaboration with the school, and include: extra time; readers and scribes; enlarged copies of the test; modified enlarged print; printed on coloured paper; braille; and access to a word processor. Schools can arrange for a student to take the test in a separate room from other students but this is always done in the presence of NFER's test administrator. NFER reports to Ofqual on the number of students who use each access arrangement.
18. During NFER's development of test items, it engages a group of experts to consider the needs of students with various disabilities. They scrutinise items to ensure they will not unfairly disadvantage these students. The items are also reviewed to ensure they do not reflect any ethnic or religious bias.

Risk Assessment

19. Two risks relating to the NRT are included in Ofqual Strategic Risk Register, 18a (NRT – delivery) and 18b (NRT – GCSE impact). In addition, the project maintains a more detailed risk register which the Project Board keeps under review.
20. This section has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

This section has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

OFFICIAL – SENSITIVE

This section has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

OFFICIAL – SENSITIVE

This section has been redacted, as its publication would be prejudicial to the effective conduct of public affairs.

21. In December 2016, Ofqual’s internal auditor conducted an audit of the NRT governance and contract management. It gave an audit rating of substantial assurance.

Regulatory Impact Assessment

22. Ofqual is working with the exam boards to determine how they will use NRT data in GCSE awarding. This ensures that proposed approach to using NRT evidence in awarding (as set out in Annex C) are proportionate and effective.
23. The exam boards have no delivery role in the NRT.

Communications

24. Information continues to be available on GOV.UK, including an overview of the NRT¹. The statutory guidance to schools will be updated in September. The Chief Regulator’s letter to schools in June included a short section on the NRT.
25. In the short-term, our communications are designed to ensure sampled schools agree to take part. We are also managing expectation that

¹ <https://www.gov.uk/government/collections/national-reference-test-information>

OFFICIAL – SENSITIVE

results from the NRT will not be used before 2019 at the earliest. The NRT is covered in regular meetings with stakeholders.

26. We will develop a more detailed communications plan to explain to stakeholders how the results from the NRT will be used in GCSE awarding but we first need to agree the approach with the exam boards.

Paper to be published	<i>YES (except three closed paragraphs)</i>
Publication date (if relevant)	At later date, when Ofqual publishes information on the NRT 2017
If it is proposed not to publish the paper or to not publish in full please outline the reasons why with reference to the exemptions available under the Freedom of Information Act (FOIA), please include references to specific paragraphs	<i>Paragraphs indicated in report to remain closed. Publication would be prejudicial to commercial interests (section 43 of FOIA)</i>

ANNEXES LIST:-

ANNEX A SAG's advice on the 2017 NRT

ANNEX B Summary of National Reference Test in 2017

ANNEX C Proposal for the use of NRT outcomes in GCSE awarding (discussed with the Board in May 2016)

Annex A

Advice from the Standards Advisory Group

The NRT Sub-Group of SAG met on 22 June and provided the following advice to the Standards Advisory Group. SAG endorsed the advice at its meeting on 30 June:

The SAG NRT Sub-Group has concluded that the quality of the National Reference Test is very good. The marking has been well organised and the administration in the schools has gone very well. During the meeting, the Group offered the following advice:

- 1. The level of precision of the test achieved in 2017, if maintained at a similar level in subsequent years, is appropriate for the results to be taken into account in the awarding of GCSEs.**
- 2. There should be an evaluation of the scale of the benefits from using complex IRT models compared with more conventional equating and estimation of population outcomes.**
That work should help us understand whether more conventional approaches would produce data that were adequate when using the NRT outcomes in GCSE awarding.
- 3. The current sampling frame should be maintained.**
Independent schools are not subject to the legislation and were much less willing to take part in 2017 than state schools. Although this may have resulted in some bias in the NRT results, it is believed to be modest. Retaining independent schools in the sample to maintain face validity of the test is considered preferable. The impact of independent schools' response each year should be monitored to determine whether this might have an impact of any measure of change in student performance observed through the NRT.
- 4. It does not seem necessary or worthwhile to try to weight the student responses in the NRT using their prior KS2 results.**
KS2 scores should not be used to adjust weights when analysing the NRT test results as the takers' KS2 score distribution is sufficiently similar to that of the whole sample, and the loss of data due to non-matching may counteract any gain from using KS2 data to adjust weights.
- 5. The eight test booklets containing the 'live' items should be used to measure students' abilities.**
In 2017, the option to include responses to the items in the four 're-fresh' booklets was analysed, potentially, to achieve an improved level of test precision. The results were inconclusive and therefore there is insufficient justification to use all 12 booklets for this purpose. The NRT

OFFICIAL – SENSITIVE

outcomes should be based on the eight 'live' booklets. Each year, the NRT administration will continue to include re-fresh booklets to enable additional items to be trialled and to build up a bank of items that could be introduced into the live test in the future.

6. The student survey

The Sub-Group confirmed that Ofqual should continue to include the student survey with the NRT in 2018. Further thought should be given to whether some questions in the survey might be changed from year to year to gain more information. The analysis of the survey data should be brought to a future meeting of the Sub-Group for review.

7. Managing expectations and communications.

Ofqual should design and implement a strategy for explaining to a range of stakeholders how the NRT is designed and how its results will be used in GCSE awarding.

Annex B

Summary of National Reference Test in 2017

This annex provides a summary of the administration of the National Reference Test carried out in schools in 2017 and the results from the test.

<p>Test materials</p>	<p>The live test in 2017 used the same items and mark schemes that had been trialled successfully in the Preliminary Reference Test in 2016. In both maths and English, the live test has 200 marks of which each student takes 50 marks' worth of items, in a spiralled design of eight test versions.</p> <p>In addition, NFER had developed new items (100 marks of maths items and 100 marks of reading items for the English test) which were combined with live items in four separate test versions. Students were not aware whether they were responding to live or trial items. Items that performed well in the trial will be added to our bank of re-fresh items that could be used in the NRT in future years.</p>
<p>School sample</p>	<p>Ofqual required NFER to achieve a sample of at least 300 schools to take part in the NRT. To allow for operational problems, NFER set a target to recruit 330 schools.</p> <p>NFER also had to allow for some schools continuing to enter students only for the international GCSE, which would make the school ineligible to take part in the NRT for that subject. Schools' entry policies could only be determined when NFER contacted the selected schools.</p> <p>NFER prepared a stratified sample which included 360 schools in the 'main' sample. For each school in the main sample there were up to 3 replacement schools chosen in case the selected school was not prepared to take part. Schools' mean GCSE grade in 2016 was used to construct the strata. All secondary schools in England were included in the population, except for very small schools and special schools. Under legislation introduced in September 2016, most state schools were required to take part in the test, if asked to do so.</p>

	<p>341 schools agreed to take part, of which 314 were from the main sample of 360 and the remaining 27 being replacement schools.</p> <p>Almost every state school that was subject to the legislation agreed to take part.</p> <p>Independent schools were less likely to take part, either because they were ineligible (entering students only for the international GCSE) or unwilling for other reasons. Of the 341 schools that took part in the NRT, 7 were independent schools.</p>												
<p>Student sample</p>	<p>NFER selected a sample of up to 30 students in year 11 at each school to take the maths test and another 30 students to take the English test. (No student was asked to take both tests.) Of the 30 students for each subject, 24 took one of the live test versions and up to 6 took one of the trial versions.</p> <p>88% of students selected to take the test actually took part. (This compares to about 78% achieved in 2016.) Ofqual has compared KS2 scores between students in the sample who took the test and those who did not. There is some bias towards the lower KS2 scores in those who did not take the test. Ofqual will repeat the analysis when students' actual GCSE grades are available.</p> <p>The legislation allows a head teacher to withdraw an individual student from taking the test. This accounted for only about 1 in 20 of the absentees, indicating heads were using this power only in exceptional circumstances.</p> <p>In summary, the numbers of students taking the test were:</p> <table border="1" data-bbox="512 1608 1353 1879"> <thead> <tr> <th></th> <th>Live booklets</th> <th>Trial booklets</th> <th>All booklets</th> </tr> </thead> <tbody> <tr> <td>Maths</td> <td>7,144</td> <td>1,762</td> <td>8,906</td> </tr> <tr> <td>English</td> <td>7,082</td> <td>1,763</td> <td>8,845</td> </tr> </tbody> </table> <p>(About 1 in 30 students in England in year 11 took a test.)</p>		Live booklets	Trial booklets	All booklets	Maths	7,144	1,762	8,906	English	7,082	1,763	8,845
	Live booklets	Trial booklets	All booklets										
Maths	7,144	1,762	8,906										
English	7,082	1,763	8,845										
<p>Test delivery in</p>	<p>The test was held in schools during a two-week period:</p>												

<p>schools</p>	<p>20th February to 3rd March 2017. Schools selected the date and time when the test would take place within this period. Some chose to hold the two subjects at the same time, others selected different times. Students had one hour to complete the test.</p> <p>NFER assigned one or two test administrators to each school to make arrangements and to go to the school on the day of the test to administer it. NFER trained all the test administrators in January at face-to-face events.</p> <p>The test administrators used a pre-agreed script to conduct the test to ensure consistency of approach.</p> <p>Schools were able to request access arrangements for students taking the NRT, these being the same as the student would be offered when taking the GCSE. These included additional time, use of a word processor, a scribe or reader, enlarged test materials, test materials printed on coloured paper, a modified enlarged print version and braille. Some schools requested that some of their students be accommodated in a separate room from the main group. One of NFER’s test administrators was always present in each room.</p> <p>Only one test at one school did not take place as planned, caused by adverse weather which prevented the test administrator from travelling to the school.</p> <p>Ofqual attended tests at 15 schools to observe the administration.</p>
<p>Marking</p>	<p>NFER recruited and trained a total of 110 markers (including lead markers and team leaders), about half of whom had also marked the PRT in 2016. Markers for the live test were trained face-to-face on 4th March and completed the marking of the live booklets on schedule on 18th March.</p> <p>NFER controlled the quality of marking through standardisation, seeds (a rate of 1 in 20) and a further review by team leaders of 4 responses in each batch of 20.</p> <p>Extended response items in the English test were 100% double-marked with out-of-tolerance items being blind-</p>

	<p>marked by a third marker (the team leader).</p> <p>Items in the four trial booklets were marked between 19th March and 16th April.</p>
<p>Test security</p>	<p>The test items are re-used each year so it is essential we maintain the security of the items and mark schemes.</p> <p>NFER uses a secure courier to transport printed materials between its office and test administrators and markers. All items are closely tracked to ensure none is lost. One package being returned by a marker to NFER went astray whilst in the possession of the courier. After extensive investigation, the courier found the package one month later. There was no evidence that the security of test materials had been breached.</p> <p>NFER’s test administrators control all test materials during the test taking place in schools. School staff are not permitted to have access to them.</p> <p>Ofqual and NFER monitor social media to identify if any test content may have been published. There is no evidence this happened during the NRT in 2017 (or, previously during the PRT in 2016).</p>
<p>Psychometric analysis and test outcomes</p>	<p>CITO completed the analysis of the test data on schedule, using the agreed psychometrics framework. For this analysis, Ofqual provided NFER with assumed GCSE percentiles to use for the key grade boundaries. CITO will update the analysis in the autumn, once the actual percentiles are known from this year’s GCSE awards.</p> <p>We use this year’s GCSE results to establish the positions on the NRT ability scale that represent the GCSE grade boundaries. In future years, we use the NRT results to estimate the percentage of students at these positions on the ability scale from which we can see whether there may have been a change in performance to take into account when GCSEs are awarded.</p> <p>We designed the NRT with the aim to achieve a level of test precision at three specific grade boundaries of $\pm 1.5\%$ (with 95% confidence). The test precision for the NRT in</p>

	<p>2017 was:</p> <table border="1" data-bbox="512 264 1276 568"> <thead> <tr> <th data-bbox="512 264 898 338"></th> <th data-bbox="898 264 1086 338">English</th> <th data-bbox="1086 264 1276 338">Maths</th> </tr> </thead> <tbody> <tr> <td data-bbox="512 338 898 416">Grade boundary 6 / 7</td> <td data-bbox="898 338 1086 416">±1.3%</td> <td data-bbox="1086 338 1276 416">±1.3%</td> </tr> <tr> <td data-bbox="512 416 898 495">Grade boundary 4 / 5</td> <td data-bbox="898 416 1086 495">±1.9%</td> <td data-bbox="1086 416 1276 495">±1.6%</td> </tr> <tr> <td data-bbox="512 495 898 568">Grade boundary 3 / 4</td> <td data-bbox="898 495 1086 568">±1.6%</td> <td data-bbox="1086 495 1276 568">±1.5%</td> </tr> </tbody> </table> <p>Ofqual, with the advice from the NRT Sub-Group of the Standards Advisory Group, has concluded that the level of test precision achieved in 2017, if maintained at a similar level in subsequent years, is appropriate for the results to be taken into account in the awarding of GCSEs. (It should be noted that, currently in awarding GCSEs, the key boundaries are 3/4 and 6/7. Grade boundary 4/5 is determined arithmetically.)</p>		English	Maths	Grade boundary 6 / 7	±1.3%	±1.3%	Grade boundary 4 / 5	±1.9%	±1.6%	Grade boundary 3 / 4	±1.6%	±1.5%
	English	Maths											
Grade boundary 6 / 7	±1.3%	±1.3%											
Grade boundary 4 / 5	±1.9%	±1.6%											
Grade boundary 3 / 4	±1.6%	±1.5%											
<p>Student and school survey</p>	<p>Students taking the NRT were also asked to complete a short questionnaire. Students completed the questionnaire in about 5 minutes after the test. Ofqual will use this information to monitor, for example, whether there may be any changes in the motivation of students who complete the NRT. This would be contextual information that would be taken into account when considering how information from the NRT should be used in the award of GCSEs.</p> <p>Ofqual also collected information from schools via NFER’s test administrators.</p> <p>The analysis of the data from this year’s student and school surveys will be completed after GCSE results are available.</p>												

ANNEX C

**Proposal for the use of NRT outcomes in GCSE awarding
(provided to the Board in May 2016)**

This is a high level proposal regarding how we could approach determining the application of the NRT outcomes to GCSE awards in England. Detailed governance arrangement including detail of the decision making processes would have to be developed and documented.

1. At the start of May 2019², Ofqual will receive from NFER the draft technical report on the 2018 NRTs, providing statistical analyses of NRT data and their implications. It will include information around the confidence intervals around the outcomes as well as look-up tables giving equipercentile mapping of the NRT to GCSE. This report would be shared in confidence as soon as possible with the responsible officers of AQA, OCR, Pearson and WJEC³.
2. In mid-May, Ofqual would convene a technical panel to scrutinise the report. The panel would have available NFER's draft technical report from that year's NRT. The terms of reference of the panel would focus on it providing advice to the Grade Standards Forum (see step 3 below) on what actions to take in that summer's GCSE awards on the basis of the NFER report on the NRTs and other pertinent evidence such as research into the Sawtooth Effect. The membership would be:
 - Chair: Executive Director of General Qualifications, Ofqual
 - Two Ofqual representatives from Strategy, Risk and Research
 - One technical representative from each of AQA, OCR, Pearson and WJEC.
 - Two external, independent representatives appointed by Ofqual. These might be members of Ofqual's Standards Advisory Group (SAG), or from the NRT sub group of SAG that we have established.
 - NFER would be invited to send one or two technical advisers. The timing of the forum allows for NFER to be asked to provide additional analysis of the NRT results before finalising its Technical Report by the end of May.
3. In early June, Ofqual would convene the Grade Standards Forum, chaired by the Chief Regulator and including senior members of Ofqual staff together with the responsible officers of AQA, OCR, Pearson and

² In July 2016, the Board agreed that the first year when information from the NRT could be considered when GCSEs are awarded will be 2019. The proposal set out in this appendix has been updated to reflect this decision. The version discussed with the Board in May 2016 referred to 2018 rather than 2019.

^{3 3} We plan a 'dry-run' of the process in 2018 so the boards will be clear what to expect in summer 2019.

OFFICIAL – SENSITIVE

WJEC (plus their technical experts). To help logistics, the meeting might take place on the morning of the June Maintenance of Standards meeting.

The purpose of the Forum would be to discuss what action would be appropriate to take in that summer's GCSE awards on the basis of the advice from the technical panel. This would be informed by the advice from the technical panel. The decision would be Ofqual's to take. The role of the boards at the Forum is so that they can make representations. There might also be benefits in inviting observers, for example, from the head teacher and teacher associations, and from the Wales and Northern Ireland regulators but their status and purpose would have to be made very clear.

4. Following the Forum, Ofqual would have to make a regulatory decision about exactly how that year's NRT outcomes would be applied in that summer's GCSE awards. The process for that decision will have to be carefully developed, agreed and documented.
5. The actions would then be built into Ofqual's data exchange processes, to be issued to the boards in the second half of June with the NRT-related part not published before GCSE results day. (NFER is contracted to provide a General Report of that year's NRT results that is suitable for publication. These data will be Official Statistics.). Additional conditions of recognition or regulatory requirements underpinned by conditions might also be required.
6. Assuming the process above described is successful, it would be repeated on an annual basis.