

# Ofqual Board

## Paper 68/14

**Date:**

3 December, 2014

**Title:**

National Reference Tests

**Report by:**

Tim Leslie, National Reference Test Project Director

**Responsible Director:**

Michelle Meadows, Acting Executive Director for Strategy, Risk and Research  
Marc Baker, Acting Executive Director General Qualifications

**Paper for decision**

**Closed (to be published after the contract has been awarded)**



**Issue**

1. In Ofqual's Corporate Plan 2014 – 17, we confirmed our plan to introduce a National Reference Test to strengthen the awarding of GCSEs. We have developed our requirements for the new Tests and issued an Information to Tender (ITT). Suppliers had until 21 November to respond to the ITT and we plan to award the contract before end January 2015. This paper sets out our proposals for the Tests and how we are conducting the procurement, and seeks Board approval to proceed to award the contract.

**Recommendation**

2. The Board is invited to:
  - a. Approve that Ofqual proceeds to invest in the development of The National Reference Tests for the purpose stated in the Information to Tender, issued on 26 September 2014, as set out in paragraph 5 of this paper, subject to the outcome of the evaluation of responses received from potential suppliers.

## OFFICIAL – SENSITIVE

- b. Delegate to a Committee of the Board to approve Ofqual entering into an Agreement with the Preferred Supplier to provide the Solution, the Preferred Supplier having been determined through the evaluation of the tenders that have been received, subject to the Committee:
  - i. scrutinising the procurement process that was followed, including the evaluation of the tender responses, to be satisfied in the integrity and robustness of its outcome
  - ii. reviewing an updated assessment of the risks to Ofqual from awarding the Agreement, the risks having been updated after the tender evaluation is complete, and confirmation that DfE has committed sufficient funds until March 2016.
  - iii. confirming that the Preferred Supplier has the capability and capacity to deliver Ofqual's requirements for the Tests for the term of the Agreement.

### **The Board's previous considerations**

3. The Board was advised in January 2014 (paper 103/13 Setting GCSE and A level Grade Standards), of the following rationale for the introduction of a Reference Test.
  - a. While the use of comparable outcomes at A level has been little criticised of late, the same cannot be said for GCSE. This seems to have been a consequence of the different contexts within which these qualifications operate and their different purposes.
  - b. A level results are primarily used for selection into higher education courses. From the universities' perspective, keeping the national A level grade outcomes broadly constant from year to year serves them well.
  - c. At GCSE the position is different. Schools in the state sector feel under great pressure from the Government's targets, particularly expectations that proportions of 16 year olds having achieved grade Cs in high profile subjects will rise year on year. There are currently no similar pressures for schools and colleges in relation to 18 year olds. A clear tension has arisen between Government expectations for 16 year olds' attainment and the application of the comparable outcomes approach at GCSE beyond the first year of new exams. The implication of keeping the comparable outcomes approach in years 2, 3, 4 etc of the GCSE exams is that national grade C outcomes will remain broadly constant from year to year despite schools' increasing efforts to improve their performances.

## OFFICIAL – SENSITIVE

- d. The Ofqual public position has not been that national grade C outcomes will necessarily remain the same from year to year. We say on our website:

*We believe that grade inflation – year-on-year increases in results without any real evidence of improvement in performance – should be avoided. It undermines confidence in the qualifications and in students' achievements. Our approach aims to control grade inflation, but to allow genuine improvements in performance to be recognised.*

- e. The problem lies in how the comparable outcomes approach squares with allowing “genuine improvements in performance to be recognised”. This is not an issue about the first year of new exams. It is in the use of comparable outcomes in the following years.
- f. If there is a genuine improvement in performance of students in the second year of an exam it is likely that is largely because their teachers are more familiar with the requirements of the course and the nature of the exams and so are better able to prepare students. It is unlikely that this improved performance indicates that the latter cohort is substantially better in terms of, for example, their capacity for future learning. If we don't want unfairly to advantage the second cohort, the use of comparable outcomes appears appropriate. In doing so we should acknowledge that any small increase in, for example, students' capacity for future learning would not be recognised by increases in greater proportions of higher grades.
- g. Suppose though that in the fourth year of a GCSE examination there are *genuine improvements in performance* of the students. The Ofqual position is that this can be an acceptable justification for the proportion of students awarded a higher grade that year to rise.
- h. The challenge arises from the nature of that evidence we require. Using only their judgement of scripts, awarders are unable to make the fine judgements necessary to decide whether a grade boundary should be put at, for example, 62, 63 or 64 marks. If that is the case, without an improbable great hike in performance from one year to the next, it is demanding for awarders' judgements to provide persuasive evidence. Indeed that is the justification for the use of a reference test to help us maintain grade standards – in a performance sense – in the new GCSEs.
- i. As the Chief Regulator said to the then Secretary of State in her 22 August 2012 letter, our comparable outcomes approach can make it harder for genuine improvements in performance to be

## OFFICIAL – SENSITIVE

fully reflected in the results. It is important though that we remain open to the possibility that an exam board could present us with evidence in this regard which, after careful consideration, we concluded did indeed justify an out of tolerance award.

- j. The challenge described above is the justification for the use of National Reference Tests to help us maintain grade standards – in a performance sense – in the new GCSEs. The purpose of the Reference Tests is to provide Ofqual and the exam boards with that credible, persuasive evidence that will allow us to justify a change from one year to another in the proportion of 16 year olds getting a high grade in their GCSEs.
4. The Board had previously been advised at its meeting in July 2013 that DfE had consulted on the introduction of a national sample test as part of its proposals for changes to the school accountability measures at Key Stage 4. Following discussions between Ofqual and DfE, ministers had indicated that Ofqual should own a more focused reference test to help set and maintain standards for new GCSEs and that funding would be provided for this work. The Board noted that the introduction and operation of the test would be a substantial piece of work. In reaching this position, Ofqual had considered other options that might support the maintenance of standards in the new GCSEs but concluded that none offered a more realistic prospect of meeting this need. For example, common items could be embedded in each exam board's GCSE papers each year. It would be far harder to protect the security of these items and would require considerable coordination with all the exam boards.
5. In April 2014, Ofqual included outline proposals for a reference test in its consultation on Setting Grade Standards for new GCSEs in England. The Board noted the outcome of the consultation at its meeting on 19 August and were advised that more detailed proposals had been included in a draft ITT which had confirmed the purpose of the national reference tests. **The purpose is to provide evidence for Ofqual on changes in performance standards over time in GCSE English language and mathematics in England at the end of Key Stage 4.**
6. Our aim is to create Tests that provide an anchor for GCSE standards. This is the primary purpose of the Tests. We have not proposed tests that measure a wider curriculum because then they would not serve the anchor function as well. The Tests are being designed to be at a level of generality that will make it robust against low level changes to the GCSE curriculum and resilient to teaching to the test (as far as a supplier can design it, given that it has to be an anchor).
7. Although they will enable us to track national performance in English language and maths at whole subject level, the Board should note that the stated purpose does not include monitoring of students'

## OFFICIAL – SENSITIVE

performance in specific areas of the curriculum. For example, the Tests would not reliably tell us about how the performance of England's 16 year olds is changing year on year in using language creatively in writing or using algebra to create proofs.

8. The purpose also makes explicit that:
  - a. The Tests will be developed for use in England
  - b. They will be set only in the subjects of Maths and English Language
  - c. They will measure changes in performance standards of students at the end of Key Stage 4, therefore not taking into account changes in performance of other candidates that may also take these GCSEs.

### **Using the Tests in the award of the new GCSEs**

9. The Tests will provide additional information to support the awarding of GCSEs in Maths and English Language. We will need to consider further how the current limitations of comparable outcomes might be addressed in awarding of new GCSEs in other subjects. The information from the Tests will not replace any information that is currently used. So the statistical information that is currently used, including the Key Stage 2 predication matrices, combined with examiner judgement will remain at the core of each exam board's awarding process.
10. The additional information from the Tests will be used in the early stages of each year's awarding cycle to determine whether, for the national cohort, there should be an expectation that the proportions of students in each grade should be adjusted to reflect a change in performance standards. We have not designed the Tests to be used in our work, ensuring inter-board comparability, for which we will continue to use existing practices. After we have awarded the National Reference Test contract, we will work with exam boards to develop the approach in more detail and will bring these proposals to the Board for approval.
11. The first cohort of students to take the new Maths and English Language GCSEs will sit the exams in summer 2017. This will also be the first year when students will take the fully developed National Reference Tests. (We plan to carry out a Preliminary Reference Test in 2016 – see paragraph 22 for further details.) In 2017 after the GCSEs have been awarded, we will use the proportions of students in each grade in the GCSEs to establish the marks in the National Reference Tests that correspond to performance standards at the key grade boundaries of 3/4, 4/5 and 6/7. As an illustrative example, say 55% of students achieve a grade 4 or better in a GCSE in 2017. In the Reference Test, 55% of a representative sample of those students

achieved, say, at least 38 marks. This would establish the minimum mark in the Reference Test representing the performance standard for grade 4 to be 38.

12. In subsequent years, we will know how students at the end of Key Stage 4 performed in the Tests, before GCSEs are awarded. Using the same illustrative example, we might see that 57% of students achieved a mark of 38 or better, a 2 percentage point increase in students' performance in the Test in 2017. This might therefore indicate that 57% of students at the end of Key Stage 4 might also be awarded at least a grade 4 in the GCSE. The information from the Reference Tests will be only one source amongst several that, used together with other statistical data and examiners' judgement, will determine the awards of the GCSEs.

### **Outline design of the National Reference Tests**

13. In the Invitation to Tender, we have set out the high-level design of the Tests. It will be the Supplier's responsibility to develop a more detailed design, based on our requirements.
14. The content domains for the Tests are to be the same as the subject content that DfE specified for the new GCSEs in Maths and English Language. The Maths Test is to be based on the GCSE higher tier content. The English Language Test will include both critical reading and comprehension, and writing, but will not include spoken language. The cognitive domains for the Tests are to be the same as set out in the DfE's associated Assessment Objectives.<sup>1</sup>
15. While the design of the Tests will therefore mirror the GCSEs in Maths and English Language, we want to distinguish genuine changes in performance from changes that are due to increasing familiarity with the typical content and style of GCSE questions or that do not generalise to the domains of Maths and English Language. We have therefore specified that questions to be used in the Tests, while being accessible to students preparing to take their GCSEs, are not too close in style to those used in the GCSE assessments. We therefore aim to make the Tests to be resistant to the possible increase over time in formulaic approaches to question answering.
16. The Tests will take the form of overlapping booklets, where the same question is used in more than one, but not all the booklets. There will be eight overlapping booklets in each of the Maths and English Tests. Each student that takes a Test will take one booklet in either Maths or English. Therefore no student responds to questions covering the

---

<sup>1</sup> <https://www.gov.uk/government/publications/gcse-english-language-and-gcse-english-literature-new-content>

<https://www.gov.uk/government/publications/gcse-mathematics-subject-content-and-assessment-objectives>

## OFFICIAL – SENSITIVE

complete curriculum. There will be no results provided to individual students or aggregated for all the students taking a Test at an individual school. The results will be combined using a psychometric framework to derive a prediction of the distribution of marks in the Test across the national cohort. We have asked the bidders to decide whether to require students to complete the Tests on paper or on screen. The same questions will be used each year.

17. We have determined the level of precision we need in the Tests' results for their use in GCSE awarding whilst keeping the sample sizes of those taking the Tests to be operationally feasible to deliver. Ofqual has carried out detailed modelling of options for the level of precision, sample sizes and structures and has specified its requirements in the Information to Tender as:

- A precision of +/- 1.5% at the 95% confidence interval at the three prescribed grade boundaries (grades 3/4, 4/5 and 6/7)
- A two-stage sampling model will be used; first, the schools to be sampled are to be selected, stratified according to their GCSE performance. Students are then sampled at random from the selected schools. Different schools will be selected each year to take part in the Tests. The Reference Test Supplier is responsible for selecting the schools and students.
- A sample of 300 schools will take part in each of the Maths and English tests each year; the Supplier can decide whether to use the same schools for both subjects or to use different schools for each.
- A sample of 25 students at each school will take the Test, providing a total of 7,500 students for each of the Maths and English tests.

18. Bidders that respond to the Initiation to Tender can offer alternative sample sizes for both the number of schools and number of students but must demonstrate that their solution will achieve the required level of precision in the Tests' results.

19. The Supplier will be responsible for recruiting the schools to take part in the Tests each year. However, it is recognised that some schools may be reluctant to take part. After the contract has been awarded, Ofqual will develop a campaign to engage with schools to encourage their participation. Additionally, Ministers have agreed in principle to introduce secondary legislation to require a school, if selected, to take part. DfE officials are drafting this legislation which will be based on existing legislation that requires schools to participate in the Key Stage 2 science sampling tests. We have requested the secondary legislation to be in force from the start of the academic year 2015-16. We have proposed that the Tests are held each year in early March so that they do not interfere with students' final preparations for their GCSE exams.

## OFFICIAL – SENSITIVE

20. The Reference Test Supplier will be responsible for conducting the Tests in schools. Schools therefore will need to provide suitable accommodation for the Tests and to ensure that the students selected to take the Tests are available. But it will be the Supplier's own resources that will administer and invigilate the Tests. We have proposed this approach both to minimise the disruption for schools and to protect the confidentiality of the Tests' questions.
21. Following completion of the Tests, the Supplier will be responsible for marking the students' responses and then to derive a measure of performance for each of Maths and English for the national cohort. We require the completed analysis of the Tests' results to be available by end May, before the start of the GCSE awarding process.

### **National Reference Test Agreement – scope of services**

22. The Information to Tender includes the legal Agreement that Ofqual would enter into with the Preferred Supplier at the end of the procurement. We considered the option to procure a supplier to design and develop the Tests first, then to procure based on that design, a supplier to deliver the Tests. We decided against this two-stage option because there are relatively few organisations in this market and it would be difficult to achieve value for money on the potentially higher-value delivery contract, other suppliers being at a disadvantage to the Tests' developer. The Agreement sets out three Services:
  - Strategic Design Services: the initial phase of work during which the Supplier will prepare a high-level design for all aspects of the Tests – we anticipate this will last about 4 to 6 months
  - Detailed Design and Development Services: following on from the previous phase, the Supplier will complete the development of the Tests. This includes writing the Test questions and mark schemes. We require the Supplier to carry out a Preliminary Reference Test in March 2016. This will be a full-scale trial. Students in year 11 that take the Tests in 2016 will be preparing to take the current GCSEs and therefore results from the Tests cannot be used to establish the performance levels in the Tests that equate to grades in GCSEs. The primary purpose of the 2016 Preliminary Reference Test is to confirm all aspects of the Tests are operating satisfactorily. However, it may be possible in 2017 to derive some information by comparing responses to individual questions across the two years. This may assist in the first award of the new GCSEs in 2017. The Supplier will make any changes to the Tests and how they are delivered following feedback from how the Preliminary Reference Test has performed. The Detailed Design and Development phase will end in late 2016
  - Delivery Services: the annual cycle of testing, commencing with the Tests in 2017. The Agreement includes four annual Test cycles and also gives Ofqual the option to extend the contract for a further two

years. The Term of the Agreement has been chosen to ensure that, subject to the Supplier meeting its obligations, there will be a sufficient number of annual test cycles to establish how well the Tests perform before Ofqual has to re-tender for the Services.

**Ofqual's development of its requirements and the ITT**

23. The project to establish Ofqual's requirements for the National Reference Tests has been managed as part of the Reform Programme. The Standards Directorate, and subsequently the Strategy, Risk and Research Directorate, have led the development of our technical requirements. Ofqual has also benefitted from the advice of an Expert Group (Professor Mike Cresswell, Professor Jo-Anne Baird and Colin Watson of the STA) as we developed our requirements.
24. In March, representatives from Ofqual visited Hong Kong to study how a Reference Test has been introduced there to support the reform of its senior secondary qualifications. Their findings have been taken into account in developing requirements for the National Reference Tests in England.
25. We have also discussed our requirements as they developed with prospective suppliers and exam boards. In February, we held meetings with suppliers who had expressed an interest in delivering the Tests. In June, we issued a draft ITT to give prospective suppliers an opportunity to comment on our proposals, including the draft Agreement, before they were finalised. We also held a meeting with suppliers in October after we had issued the ITT and allowed them to ask questions about our requirements until 5 November (answers being published to all suppliers that had registered their interest in the bid). Suppliers had to respond to our ITT by 8am on 21 November.
26. In addition to the ITT setting out our requirements for the Tests, it also includes the Agreement that Ofqual will enter into with our chosen Supplier. We are following the OJEU Open Procedure which does not allow for any negotiation with bidders after the ITT has been issued and we therefore had to provide in full the contractual terms we will use. The Agreement is based on the Cabinet Office's Model Services Contract, issued in March 2014 for use by Government departments and other public sector organisations, adapted to accommodate the specific requirements for the National Reference Tests. The Agreement ensures that Ofqual can transfer the Tests to a replacement supplier at the end of the first contract, retaining continuity in the design and delivery of the Tests and, therefore, comparability in results over time.

**Selecting the Preferred Supplier and evaluating the bids**

27. To comply with public sector procurement rules, we set out in the ITT how Ofqual will evaluate the bids that we receive and decide the

## OFFICIAL – SENSITIVE

Preferred Supplier to which the contract will be awarded. In summary, our approach will be:

- To reject any supplier which fails to meet certain mandatory requirements that apply to all procurements in the public sector (e.g. the Supplier or its directors have been convicted for the offences of bribery or fraud)
  - To consider whether to reject a Supplier if, for example, it has not demonstrated to Ofqual that it has the necessary financial standing that it will be capable of performing the contract; we will also consider whether the Supplier has adequate insurance cover, whether there are outstanding claims or litigation against the company, and the company has an appropriate equalities policy in place
  - To consider whether the Supplier has the capability and capacity to deliver Ofqual's requirements for developing and delivering the Tests; bidders are required to include details of other relevant contracts they have delivered, to set out their organisation's relevant technical capability and the CVs of key project personnel that will work on the contract
  - To evaluate the 'quality' of suppliers' solutions by considering their responses to 46 questions set out in the ITT; the responses from the selected Supplier will become the 'Supplier's Solution' in the Agreement and therefore binds them to deliver what was offered in the ITT response. The response to each question is evaluated by several individuals and scores are then aggregated to provide an overall score, reflecting the quality of each bid. The quality score contributes 70% towards the overall evaluation of the solution
  - The remaining 30% is allocated to the 'value for money' of each solution. We are measuring value for money by comparing the total contract value for each bid, including costs to develop the Tests and to deliver them each year including for the optional two-year contract extension to 2022
  - The scores from the 'quality' and 'value for money' evaluations are combined to rank the bids and to identify the one that has won. Ofqual is not obliged to award the contract but it cannot select any other bid in preference to the one that has won the evaluation.
28. We will also invite suppliers to present their proposals to us. We will not evaluate separately their presentations but will use the additional information we obtain in evaluating their bids. We will also take up client references to provide some assurance on the suppliers' capability and track record.

## OFFICIAL – SENSITIVE

29. The EU Open Procurement procedure does not permit Ofqual to negotiate with suppliers to amend their solution or price. However, we can issue questions to suppliers to clarify their proposals and their responses may change our evaluation of their bids.

### **Board decision and timetable**

30. At paragraph 2, we invite the Board to delegate to a committee the final decision to award the contract to the Preferred Supplier, which has been selected through the evaluation process. We have asked the Board to delegate to a committee because we would like to be able to advise bidders of the decision before the Board next meets on 28 January 2015. The suggested membership of the Committee is Amanda Spielman (to chair the Committee), Tom Taylor, Maggie Galliers and Glenys Stacey.

31. The timetable for awarding the contract is:

- Ofqual opened bids at 08.00 on 21 November
- Bids are evaluated between 24 November and 6 January; suppliers present their proposals to Ofqual in the week of 8 – 12 December
- National Reference Test Project Board confirms outcome of the evaluation and the Preferred Supplier on 8 January
- Senior Management Group agrees recommendation to the Board Committee on whether to award the contract to the Preferred Supplier on 9 January
- Board Committee meets on 14 January to make decision
- Bidders are informed on the outcome of the procurement on 16 January; EU procurement rules then require a 10 day 'stand still' period during which time any unsuccessful bidder may challenge the outcome
- The contract is awarded on 30 January.

32. Ofqual has received only one response to its ITT. This does not invalidate the procurement procedure that we are following. The bid will be evaluated as planned. When we issued the draft ITT in June, three prospective bidders provided detailed feedback through both meetings with Ofqual and in written responses. After the final ITT was issued in September, two of the three came together to consider a joint bid but on 11 November they advised Ofqual that they had withdrawn from the procurement following their own assessment of the risks.

### **Funding the development and delivery of the Tests**

33. When DfE agreed that Ofqual should take forward the development of the Reference Tests, it also agreed to provide additional funding for

## OFFICIAL – SENSITIVE

this work. DfE is providing £0.5m in 2014-15 and up to £3m in 2015-16 specifically to support the development of the Tests.

34. After the General Election in May 2015, there will be a spending review, possibly covering the following three financial years. It may also question in-year spending plans for 2015-16. Ofqual will need to set out its case for funds to be included in its settlement to support the remaining work to develop the Tests in 2015 and 2016, and to carry out the annual cycles of testing in 2017 and 2018. The proposed Agreement between Ofqual and its Supplier includes provision to terminate 'for convenience', for example, were Ofqual not to be provided with sufficient funds. The terms applying to termination are standard clauses included in the Cabinet Office's Model Services Contract and are consistent with the guidance that was recently set out in the letter to all Accounting Officers on this subject.<sup>2</sup>
35. Information on the proposed prices will be provided to the Board at its meeting on 3 December.

### **Risk assessment**

36. At its meeting on 24 October, Finance Committee reviewed the risks to Ofqual arising from the introduction and use of the National Reference Tests. **Annex A** sets out the significant risks, their evaluation and mitigating actions. The evaluation of the risks reflects the position before responses to the ITT have been received. The Board will be advised at its meeting on 3 December of any material changes to the severity of the risks resulting from the bids that have been received.
37. In the ITT, we have highlighted to bidders how the Agreement manages the allocation of some key risks between Ofqual and the Supplier. This information is provided in **Annex B** of this paper. The Board should note that Ofqual remains liable to fund any additional work that may be required were there to be a breach of security of the Test questions unless the breach was the Supplier's fault.

### **Equality Impact Assessment**

38. Our requirements set out in the ITT take into account Ofqual's Public Sector Equality Duty. In particular, the Supplier is required to:
  - minimise bias in the questions that are included in the Tests. The Supplier is required to take this requirement into account throughout the delivery of the Services
  - have in place clear arrangements for making reasonable adjustments for disabled learners

---

<sup>2</sup> <https://www.gov.uk/government/publications/dao-0214-early-termination-of-contracts>

## OFFICIAL – SENSITIVE

- monitor and be able to report on the National Reference Tests for any feature which could disadvantage a group of learners who share a particular characteristic
  - ensure that its sampling framework does not exclude students with protected characteristics from being selected to take the Tests.
39. When Ofqual consulted earlier this year on Setting Grade Standards for new GCSEs in England, which included outline proposals for the Reference Test, there were no responses that drew attention to possible impacts of the Tests on any particular protected group. The Tests will not have the same ‘high stakes’ for either individual students or schools that qualifications such as GCSEs have. Our primary focus is therefore to ensure that the Supplier makes appropriate adjustments in the design and delivery of the Tests to enable those students that have a particular protected characteristic, for example a disability, to take the Test without being disadvantaged. This will also ensure that the results from the Tests reflect the national cohort that takes the GCSEs at the end of Key Stage 4. At this stage in the development of the Tests, we have not identified any potential impact of our planned approach that would have a negative impact on students because of their disability, racial group, gender, age, religion or belief, pregnancy or maternity, sexual orientation or as a result of gender reassignment.
40. We will continue to monitor closely the potential impacts of the Tests on those with particular characteristics throughout the design, development and delivery of the Tests to ensure reasonable adjustments have been made. As noted in paragraph 37, we will require the Supplier to monitor and report to us on potential impacts of the Tests. We will use this information to meet our obligation to monitor the impacts.

### **Regulatory Impact Assessment**

41. The proposed National Reference Tests will provide additional information to support the awarding of GCSEs. The information does not replace existing sources of data or examiner judgement. We will require exam boards to continue to apply our Comparable Outcomes approach. The introduction of the Tests therefore will be an enhancement to the existing policy rather than a replacement of it.
42. We have identified the following potential impacts that would result from the introduction of the Tests:
- Schools selected to take part in the Tests will need to provide accommodation and probably some staff to oversee the Tests. However, the Supplier will be responsible for the administration and invigilation of the Tests. It is unlikely that schools will incur additional costs to enable the Tests to take place. In addition, there will be some disruption to the lesson timetable. Prior to the Tests, a member of the school’s staff will need to support the Supplier in

making the arrangements. This should not require a significant amount of time and will be spread over several months.

- Students that take the test will need to be taken out of their scheduled lessons. We have proposed that the Tests take less than one hour to complete, therefore keeping this disruption to a manageable level. They are not required to do any preparation for the Tests but the Tests take up students' time. It is estimated that the general public values their time at around £6/hour. We have proposed that about 15,000 students will take a Test each year, which equates to approximately £90,000 for the value of 'lost' time for the students. (15,000 students is approximately 2½% of the national cohort.)
  - Exam boards will need to adapt their awarding procedures to take into account the additional information that will be available from the Reference Tests. This should not pose a materially greater effort. Exam boards will also need to work with Ofqual to oversee the development and use of the Reference Tests, which will involve additional workshops and meetings, and possibly participation in more detailed research.
43. These impacts should to be compared with the potential benefits from introducing the Tests. It is not feasible to put a financial value to the additional information that the Reference Tests will provide. However, the total annual entry fees for GCSEs in maths and English Language (assuming a Key Stage 4 cohort of about 600,000 and a fee per entry of £30) are approximately £36m. It is appropriate to set the annual cost to run the Reference Tests in this context. As an illustration, were the annual cost to be £1.8m, this would add approximately 5% to the cost to the exam system by introducing the Tests for maths and English Language. (The Board will be provided with information on the annual costs, based on bids received, at its meeting on 3 December.)
44. Another approach to consider the potential value from the Tests is to consider how many students might be awarded a different grade. We have specified the minimum precision for the Tests to be +/- 1.5%. Were the Tests to indicate a smaller change in performance over time than this, it is unlikely that the Test's results would have an impact on the awarding of GCSEs. Therefore consider a scenario where the Tests do indicate that there has been a change in performance of 1.5%. This could result in about 10,000 students being awarded a different grade in either maths or English Language.

#### **Communications planning and stakeholder engagement**

45. Ofqual has set out its intention to introduce the National Reference Test in its Corporate Plan 2014-17, published on 19 August. We also included an outline of our proposals in the Setting Grade Standards for GCSEs' consultation, issued last April. The draft and final versions of the ITT have been made available to exam boards and DfE. However,

## OFFICIAL – SENSITIVE

we have not yet provided details of our plans for the Tests to a wider group of stakeholders or to the public in general. Our approach has taken into account that:

- Our requirements for the detailed design of the Tests have been developing during the year as we prepared the ITT; communication of our requirements during this formative period could have caused confusion
  - We will have greater certainty in the details of the Tests after we have chosen the Supplier
  - The first time that a significant number of students will take a Test will be in 2016 (the Preliminary Reference Test); the first time that the results from the Tests may have an impact on the awarding of GCSEs could be in 2018 (and, more likely, 2019); therefore there would be limited benefits from communicating greater detail at this early stage in the development of the Tests.
46. We have also taken care to treat all prospective bidders in the same way when we have communicated information about the Tests. In particular, we have not held more detailed discussions with the GCSE exam boards than with other bidders because, at the start of the procurement process, exam boards expressed interest in bidding for the work. This has, inevitably, limited the discussions we have had with the exam boards on how the results of the Tests will be used in the awarding of GCSEs. We shall engage with exam boards on this once we have awarded the contract for the development of the Tests.
47. Although Ofqual will be responsible for the Tests, we need the exam boards to be supportive and to contribute throughout the development of the Tests. We are developing plans to establish a forum to include representatives of the exam boards, also including independent experts, to provide oversight of the development of the Tests and, subsequently, their use in GCSE awarding. We will bring our proposals to the Board for approval.
48. We are also developing a stakeholder engagement plan to support the introduction of the Tests. A key objective of the stakeholder engagement is to build a consensus amongst schools that their participation in the Tests is worthwhile for the overall good of the sector and their students. One of the significant risks to the success of the Tests is schools not agreeing to participate. Whilst the Supplier will have the primary responsibility to recruit schools to participate in the Tests each year, Ofqual needs to build a broad level of support in the sector for the Tests.

**OFFICIAL – SENSITIVE**

Paper to be published	YES (Annex A closed)
Publication date (if relevant)	After the contract for the development and delivery of the Tests has been awarded to the Supplier, probably end January 2015
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	It is not proposed to publish Annex A as publication could be prejudicial to the effective conduct of public affairs.

**ANNEXES LIST:-**

- ANNEX A National Reference Test Risk Register (closed)**
- ANNEX B Allocation of key risks between Ofqual and the Supplier (extract from the ITT)**

**Annex B Allocation of key risks between Ofqual and the Supplier (extract from the ITT)**

Nature of Risk	Explanation of Risk	Who bears risk?	How dealt with in the Agreement?
<p><b>Meeting Ofqual's Requirements for the NRT, including meeting the required level of precision</b></p>	<p>Ofqual has a number of key objectives for the National Reference Tests, including a specific requirement for the NRT results to achieve a precision of +/- 1.5% at the 95% confidence interval at the key grade boundaries.</p> <p>What if this is not achieved or Ofqual's requirements are not met?</p>	<p>Supplier</p>	<ul style="list-style-type: none"> <li>• The Services Description (Section 3/Schedule 2.1) sets out Ofqual's requirements for the NRT, including a statement that the Tests must deliver a precision of +/- 1.5% at the 95% confidence interval at the key grade boundaries.</li> <li>• The Services Description also sets out values for certain key attributes (e.g. sample size). (See also paragraph 5.5 of this Section 0 for details of the required values for those attributes in the Core Bid). Bidders are permitted to tender alternative values for one or more of those key attributes in a Variant Bid.</li> <li>• The Supplier is required (see clause 5.2) to ensure that the Services comply in all respects with the Services Description.</li> <li>• In Section 4 /Schedule 4.1, paragraph 4.2.3.7, the Supplier is required to confirm that the Supplier Solution it is proposing (including the proposed values for the key attributes (e.g. sample size) will meet Ofqual's requirement to deliver NRT results within the required level of precision.</li> <li>• Paragraph 6.3 of Schedule 7.1 (Charges and Invoicing) and Paragraph 3.3 of Schedule 8.2 (Change Control Procedure) make it clear that the Supplier will bear the costs of any change to the Supplier Solution - including increases/amendments to the key attributes such as sample size - required in order meet the required level of precision or any other requirement in the Services Description.</li> </ul>

**OFFICIAL – SENSITIVE**

Nature of Risk	Explanation of Risk	Who bears risk?	How dealt with in the Agreement?
<b>Schools' and students' engagement</b>	Insufficient schools and students agree to participate in the NRT, so that the required sample size is not achieved.	Supplier	<ul style="list-style-type: none"> <li>• The Services Description (Schedule 2.1) sets out the required sample/cluster size</li> <li>• The Supplier is obliged to meet these requirements (clause 5.2).</li> <li>• The payments due during the Detailed Design and Development phase and the majority of the payments due during the Delivery phase are Milestone Payments, payable on 'Achievement' of certain Milestones.</li> <li>• Before each Milestone is Achieved, the Supplier has to complete certain Deliverables and these are then subjected to either an 'Assurance' or a testing process (see Schedule 6.2 – Testing and Assurance).</li> <li>• A key Milestone in each of these phases (Milestone 2.5 in the Detailed Design and Development phase and Milestone 3.2 in the Delivery phase) will only be Achieved – and therefore the Milestone Payment payable – once the Supplier has delivered Reports which show that it has met its sampling requirements and has delivered Test results which are within the required level of precision.</li> </ul>
<b>Late delivery of Preliminary NRT results/Reports in 2016</b>	Results/Reports delivered late	Shared	<ul style="list-style-type: none"> <li>• As explained above, the Supplier will not get paid for Milestone 2.4 until it has formally 'Achieved' that Milestone and all the related Deliverables have been certified as having been Assured or Tested.</li> <li>• In addition, if the Supplier becomes aware that it will not meet a Milestone Date and there is a delay, the Supplier is required to implement an agreed Rectification Plan and must work to rectify the delay at its cost.</li> <li>• However, even if the Milestone is Achieved after the Milestone Date, the full Milestone Payment will still be due.</li> </ul>

**OFFICIAL – SENSITIVE**

Nature of Risk	Explanation of Risk	Who bears risk?	How dealt with in the Agreement?
<p><b>Late delivery of NRT results/Reports in 2017 and thereafter</b></p>	<p>Results/Reports delivered late (after end of May) in any year – so can't be used for the purposes of that year's GCSE grade setting.</p>	<p>Supplier</p>	<ul style="list-style-type: none"> <li>• See paragraphs 2.4 and 2.5 of Schedule 7.1 (Charges and Invoicing) for detail of the provisions which will apply in relation to the 2017 Test cycle and thereafter.</li> <li>• If Milestone 3.2 is not Achieved by 31 May in any year, then (in addition to the obligations set out above in relation to the implementation of a Rectification Plan) the 60% Milestone Payment attached to that Milestone will be reduced as follows:               <ul style="list-style-type: none"> <li>(a) If the Milestone is Achieved on or after 1<sup>st</sup> June, then the relevant Milestone Payment will be reduced by 10% of the total annual Milestone Payments (i.e. to 50%);</li> <li>(b) If the Milestone is Achieved on or after 15 June, then there will be a further 10% reduction (i.e. the Milestone Payment becomes 40%, not 60%)</li> </ul> </li> <li>• If the Supplier fails to Achieve Milestone 3.2 by 30 June, Ofqual will have the right to terminate the Agreement for the Supplier's breach (see definition of Supplier Termination Event).</li> </ul>
<p><b>Disclosure of Test questions</b></p>	<p>The efficacy of the NRT depends on the substantial majority of the Test questions being the same each year. If Test questions are disclosed, and have to be replaced, there may be a loss of inter-year comparability in the results of the NRT. There will also be costs incurred in rewriting questions.</p>	<p>Shared</p>	<ul style="list-style-type: none"> <li>• The Supplier will be liable for any costs resulting from a disclosure of Test questions caused by an act or omission of the Supplier, its employees, agents or contractors. This includes the costs of rewriting the questions, rectifying the reason for the disclosure and also costs of any extra testing, including field trials, required to demonstrate comparability of the new Test with the old (disclosed) Test. See paragraph 6.4 of Schedule 7.1 (Charges and Invoicing).</li> <li>• If there is a material disclosure of the Test questions which is caused by an act or omission of the Supplier, its employees, agents or contractors, Ofqual is entitled to terminate the Agreement for Supplier breach. See definition of 'Supplier Termination Event'</li> <li>• Ofqual takes responsibility for any additional costs resulting from a disclosure/breach of the security of the Test Booklets, Test Questions and/or related marking schemes which occur <u>other than</u> as a result of an act or omission of the Supplier, its employees, agents or contractors (see paragraph 6.5) of Schedule 7.1 (Charges and Invoicing))</li> </ul>