

# National assessments regulation

Annual report 2017



December 2017

Ofqual/17/6323

# Contents

- Executive summary .....2
- Introduction .....4
  - About national assessments regulation.....4
  - Context for 2017 .....4
- Section A: Priorities for 2017 .....5
- Section B: Monitoring assessments in 2017 .....6
  - Test development.....6
  - Standards maintenance .....7
  - Marking .....7
  - Results and the ‘saw-tooth’ effect.....10
  - Security of confidential assessment materials .....11
  - Moderation.....12
- Section D: Regulatory framework consultation .....15
- Section E: Future assessment arrangements.....16
- Looking forward .....18

## Executive summary

We are pleased to present this report about our regulation of national assessments in 2017. We have continued to take a risk-based approach, maintaining our focus on key aspects of validity and the highest-stakes assessments, in particular, those at key stage 2 which inform both attainment and progress measures for schools. We have met the commitments made in our corporate plan for 2017/18: to publish our content validation study, to monitor standards and risks and to provide advice to the Department for Education on proposed changes to primary assessment. We also carried out research into the 2016 key stage 2 reading test and our research and monitoring of moderation of key stage 2 writing teacher assessments is ongoing.

We are confident that our work this year has contributed to meeting our statutory objectives to promote standards and confidence in national assessments. We have provided assurance that key elements of the testing process are valid and robust. We have also identified where improvements can be made, but have not had cause to inform the Secretary of State of any significant failings in arrangements.

2017 saw the first series of key stage tests in which the new standards, set in 2016, needed to be maintained. We are satisfied that the Standards and Testing Agency (STA) took an appropriate approach to making sure that the new performance standards were effectively maintained for the 2017 tests. And our analysis of marking data provides evidence to support the conclusion that quality assurance measures for key stage 2 marking were effective.

We carried out a comprehensive evaluation of the STA's approach to sampling from the national curriculum. We concluded that the approach taken is robust and compares favourably to similar assessments internationally. Following on from this study, we also reviewed evidence about the accessibility of the 2016 key stage 2 reading test. This suggested areas that could benefit from further consideration and the STA has committed to taking appropriate steps to investigate and respond to the questions our research posed.

This year we have also reviewed and launched a consultation on our regulatory framework for national assessments. The framework sets out our expectations of bodies responsible for developing and delivering national assessments and explains our approach to regulation. Our review aims to bring our framework up to date, make our work more transparent and maintain our focus on assessment validity. We welcome all responses to our [consultation](#), which is due to close on 20 December 2017.

During 2017, we also responded to the government's primary assessment consultation. Decisions arising from this consultation will result in a number of

significant changes to primary assessment over coming years. While the focus of our regulation may change in response to events or new information, our priorities in the coming period will be informed by these decisions and by stakeholder insights, and will include:

- monitoring the development of new assessments such as the reception baseline and multiplication tables check
- continuing to focus on areas of risk, such as high-stakes teacher assessments
- continuing to monitor key technical processes that support and underpin the validity of high stakes testing.

We hope this report provides helpful insights to those interested in national assessments.

# Introduction

## About national assessments regulation

Ofqual regulates statutory early years foundation stage profile assessments and statutory national curriculum assessments (some of which are also known as 'SATs'), which together we refer to as 'national assessments'.

Ofqual's national assessment objectives, duties and powers are set out in law. We are responsible to Parliament, primarily via the Education Select Committee, rather than to government ministers. Our objectives are to promote standards and confidence in national assessment arrangements and our primary duty is to keep all aspects of national assessments under review. We focus on validity, that is, the quality of assessment. We also have a duty to report to the Secretary of State if we believe there is, or is likely to be, a significant failing in national assessment arrangements.

We fulfil our objectives primarily by observing, scrutinising and reporting on key aspects of assessment validity. We take a risk-based approach, which includes focusing on those assessments which have the 'highest-stakes', such as those relied upon within school accountability measures. As well as identifying risks to validity that can be addressed by responsible bodies to improve the quality of assessments over time, our regulation also seeks to provide independent assurance as to whether evidence suggests that processes are robust.

Ofqual can provide advice to support government decisions about future assessments, but we do not decide what national assessments there should be; nor are we responsible for the curriculum or school accountability policy. These things are determined by the Secretary of State for Education.

The primary body responsible for national curriculum assessments is the Standards and Testing Agency (STA). The STA is an executive agency within the Department for Education (DfE) and may contract with suppliers to help develop, deliver or monitor national assessments. Other organisations also have responsibilities for aspects of national assessments, including local authorities, schools and other parts of DfE, for example, teams responsible for early years assessment.

## Context for 2017

2017 was the second year of a new suite of assessments in mathematics and English. These assessments were set at a new, more demanding standard, based on the new primary national curriculum introduced in 2014. They include both teacher assessments and tests at key stages 1 (KS1) and 2 (KS2).

There was no KS2 science sample test taken in 2017. This test is administered biennially by the STA in a selection of schools to inform a national view of standards in primary science. Individual results are not provided to schools or pupils. The science sample test is next due to be administered in 2018.

## Section A: Priorities for 2017

During the 2017 assessment cycle, we continued to focus our monitoring on procedures that are critical to supporting valid test outcomes, including standards maintenance and marking processes. We also continued to focus on KS2 assessments, as the highest-stakes national assessments that form the basis of progress and attainment school accountability measures. In response to concerns from some stakeholders about the consistency of moderation of KS2 writing assessment, we decided to monitor this area and observe local authority moderation.

We engaged with the Department for Education's primary assessment consultations, providing technical advice to inform their decisions about the future approach to assessment. We published our [response](#) to the primary assessment consultation in June.

In October, we completed and published our [research](#) into the STA's approach to curriculum sampling for the new suite of KS2 reading and mathematics tests introduced in 2016, including a subsidiary review which looked into concerns expressed about the accessibility of the 2016 KS2 reading test.

We also reviewed our regulatory framework for national assessments, and launched a [consultation](#) on a revised framework in October. Our review aimed to bring our framework up-to-date, provide greater transparency and clarity about how we regulate and make sure our expectations of responsible bodies focus on outcomes such as validity.

This report summarises our activities and provides a view on key aspects of the validity of national assessments in 2017.

### **2017 priorities: summary**

1. Monitoring processes supporting the validity of summer 2017 tests, in particular test standards maintenance and marking (Section B)
2. Completing and publishing our research on the approach to curriculum sampling for the new suite of reading and mathematics tests at KS2 (Section C)
3. Revising and consulting on our national assessments regulatory framework (Section D)
4. Providing technical advice to inform Departmental decisions on future primary assessment arrangements (Section E)

## Section B: Monitoring assessments in 2017

### Test development

Educational assessment aims to provide useful and relevant information about what candidates know and can do. In order to achieve this, the STA has a well-developed, psychometrically-driven approach to developing tests, and this process was implemented for 2017's key stage tests. The STA's test development process takes approximately three and a half years and brings in views from experts (teachers, curriculum experts and disability experts) at three separate points in the process. The process generates a considerable amount of data relating to the performance of every item (question) in each KS1 and KS2 test, primarily through pre-testing (see below). Data from pre-testing helps STA understand how each particular item is likely to perform before it is taken, to help ensure that tests will allow for effective differentiation (discrimination) between pupils across the range of performance that needs to be assessed, including at higher and lower attainment ranges. Item level data from both pre-testing and live tests (for KS2, where data can be collected as tests are externally marked) also helps with standards maintenance.

While it is important that tests effectively function across the range of performance, the exact level of difficulty of most types of test, including those at KS1 and KS2, will vary from year to year. This is because it is very difficult to consistently find different groups of items each year that, overall, present precisely the same amount of challenge to pupils. STA takes an 'item-banking' approach to test development, which allows it to select from a range of items when constructing a test and provides it with data about the level of difficulty of each item before it is selected.

Some key stage tests in 2017 were of a very similar level of difficulty to 2016, however there were some small variations. For example, data suggests that in 2017 compared to 2016, the KS2 GPS tests contained more items at the higher attainment range, the KS1 reading test contained more questions assessing the lower attainment range and the KS2 reading test contained more items assessing the lower and middle ranges. In 2017, stakeholders did not raise concerns about the accessibility of the KS2 reading test, as they had in 2016. As a result of those concerns, STA had taken steps during the test development process to ensure that the 2017 KS2 reading test was more accessible. On the basis of lessons learned from 2016, STA selected texts and items from their 'item-bank' for the 2017 reading test that were more accessible than those in the 2016 test.

Variations such as these reflect how easy or hard pupils found a particular test compared to those in the previous year. However, they do not affect the comparability of standards across years. While it is preferable to have minimal variation to support a more consistent pupil experience, differences in test difficulty from one year to the next can be both understood and accounted for through the standards maintenance process. This process means that results reflect the attainment of pupils, rather than the level of difficulty of the particular test that was taken.

## Standards maintenance

Standards maintenance procedures differ for different types of assessment. For assessments that do not change year-on-year, such as the EYFSP, or teacher assessments against the same set of standards, there is normally no need for 'equating' to ensure that standards in each successive assessment are comparable to those from previous series.

However, where assessment tasks change each year, there is a need for specific control systems to be put in place to ensure that standards are comparable year-on-year. This is the case for KS1 and KS2 tests. STA uses a highly-developed psychometric process, known as 'equating' to support standards maintenance, so that even though the level of difficulty of a test may change from year to year, standards across years are still comparable. Equating is also used to maintain standards for the phonics screening check.

New, more demanding standards were set for the new suite of key stage tests taken from 2016. In 2016, Ofqual scrutinised the standard setting process for both KS1 and KS2 tests and concluded that STA had adopted an appropriate and professionally recognised standard-setting technique and applied this process carefully and effectively.

2017 was the first year that this new standard set in 2016 needed to be maintained. Ofqual observed standards confirmation meetings for both KS1 and KS2 and reviewed the assumptions upon which the equating model was based. We are satisfied that STA took an appropriate approach to making sure that the new standards set in 2016 were effectively maintained for the 2017 tests.

## Marking

KS1 tests are marked by teachers to inform teacher judgements, while KS2 tests are externally marked. External marking allows for a greater degree of control over marking quality and is a key process supporting the validity of KS2 testing and

### What is equating?

Tests that are made up of a new set of questions each year may vary in difficulty from year to year. Therefore, processes are needed to ensure that outcomes (e.g. 'pass', 'A\*' or '92') mean the same each year.

STA maintains standards through a statistical process called 'equating' which aligns standards in tests across years.

Each summer STA pre-tests every item (test question) that might be taken forward into a future year's live test with a minimum of 1,000 pupils who also take an 'anchor test' which is calibrated to the test standard.

By using performance on the anchor test, STA's psychometricians can compare live tests to previous tests to ensure that a score, say of 100, in one year represents the same (attainment) standard as a score of 100 in another year.

Pupils involved in pre-testing will not see any of their pre-test items in their live tests, as pre-tests are securely administered to year 6 pupils who will be in at least year 7 before any material from that pre-test can be used live testing.



providing results that can be relied on. There are a number of measures put in place to assure marking quality, including:

- Training all markers using the same script and training materials
- Requiring markers to pass a training exercise prior to live marking
- Testing accuracy during marking against 'validity items' that have already been marked by senior markers
- Maintaining a marking hierarchy to provide ongoing oversight and to ensure that items that markers are unclear about can be 'escalated' to a more senior marker
- Stopping markers from marking particular items if their marking is not of sufficient quality and remarking relevant items

This year, we observed KS2 marker training for reading, mathematics and grammar, punctuation and spelling tests. Just over 4,000 markers were trained on Saturday 13 May (immediately after KS2 test week) at 32 different locations across the country using a common script and training materials. We observed several elements of good practice in marker training, including training on items that are most complex to mark at the beginning of the day when markers were fresh, and using standardised training exercises for markers to complete, overseen by senior markers to ensure consistency of practice.

We also observed the selection of validity items for use as part of quality control during live marking for KS2 reading and mathematics. During the selection of sessions we observed, teams of senior markers, including a team member whose role was to provide quality assurance, carefully and expertly discussed the suitability of items for the task of monitoring and assuring marking quality. Items were selected to check that specific aspects of the mark scheme, marking principles and training were understood for each question in the test. For example, a wrongly spelled, but otherwise correct answer was selected as a validity item for the reading test to ensure that markers understood that misspellings should not prevent a correct response from being credited. Items were also rejected that were considered too straightforward to mark to be a useful check of marking quality.

#### **What are validity items?**

Validity items are responses that pupils have provided to questions, either during the pre-test or the live test, which demonstrate particular features (eg common errors). They are pre-marked by senior markers then put into the marking pool for live marking to check the accuracy of markers' judgements.

For KS2 marking, around one in 20 items each individual marker marks is a 'validity item'. Markers do not know which items they are.

Validity items are sometimes known as 'seeds'.

### *Concerns about Grammar, Punctuation and Spelling test item*

Following the return of results to schools on Tuesday 4 July 2017, concerns began to appear on social media about the marking of question 2 of the KS2 grammar, punctuation and spelling (GPS) test. This item required pupils to correctly position a handwritten semicolon into a sentence of typed text. In response to stakeholder concerns, we asked STA for assurance that:

1. There was appropriate redress to correct any marking errors on the 2017 test.
2. They would ascertain whether there may have been any potential issues with either the question or mark scheme construction and learn any appropriate lessons.
3. Marking requirements are made sufficiently clear to teachers to support effective pupil preparation for tests.

STA confirmed that its marking review process was in place and available to be used to correct any marking errors for this (or any other) item and that published mark schemes should have been followed. They also advised us that the item had performed effectively during pre-testing, but that they would review live data and marking review data, and consider if any lessons should be learned.

Ofqual also reviewed item functioning data, which confirms that the semicolon item performed effectively at pre-testing, but did not function as intended during live testing. It appears likely that this was because, during live marking, markers had taken a stricter interpretation of the mark scheme than had been intended, on the basis of training or guidance materials provided.<sup>1</sup> Despite this, there is no evidence that the functioning of this item had any impact on the overall functioning of the test. Indeed, while there were a higher number of changes made to this item during marking reviews compared to other items, changes to this item were made to only a tiny fraction of the more than 580,000 GPS tests that were taken.

In light of its own review, STA has committed to making marker guidance and training materials clearer for similar items in future, in particular: emphasising that markers should take a positive approach to marking; reducing the level of prescription in the relevant mark scheme; and placing a greater focus on professional marker judgement. STA has considered relevant data and drawn lessons from this experience, which should reduce the likelihood of similar concerns arising in future.

### *Marking review success criteria*

Where teachers believe there has been a marking error, they may apply to STA for a marking review. STA's policy is to correct any marking errors, including changes of

---

<sup>1</sup> To support quality assurance, it is common practice to carry out marker training and provide additional marker guidance materials (such as exemplars and training exercises) in relation to assessments such as these.

only one mark. However, the application of ‘success criteria’ allows a proportion of the cost (£9 per review) to be recouped if a review does not result in a change of three marks or more, or a change in whether or not a pupil meets the expected standard. As a result of questions raised this summer, both by Ofqual and by stakeholders, and in light of review data from 2016 and 2017 (the first two years of the new tests), STA have reviewed their marking review success criteria. From 2018, they will not charge for marking reviews where reviews result in changes of two marks or more (as well as retaining the current policy in relation to changes at the expected standard). Reducing charges in this way may make marking reviews easier to access for small schools with concerns about low level marking errors, where minimal differences in pupil scores will have more impact than in larger schools.

### *Marking consistency*

As part of Ofqual’s wider regulation of qualifications, we have developed metrics that allow us to quantify marking consistency. We analyse a significant amount of operational data to provide objective and evidence-based indications of marking quality.

Details about how the metrics are calculated are set out in Ofqual’s report ‘Marking metrics’ (2016). In brief, metrics are created from the data arising from the operational monitoring of quality of marking during live marking sessions. We assume that the most appropriate measure of consistency of marking is based on the difference between two marks given for a single response. Thus the data used is the mark – re-mark data from validity items (ie an analysis of the difference between the ‘definitive mark’ set by the senior markers and the actual mark awarded).

We have analysed marking data from 2017 reading, mathematics and grammar, punctuation and spelling tests, using our marking metrics. This has demonstrated that overall, KS2 marking is of a very high consistency with almost all markers (99.5%) agreeing with the definitive mark for 6.7 million marked items. It compares favourably to the consistency of similar (1, 2 and 3-mark) items in general qualifications (which is to be expected, given the different levels of complexity of KS2 and GCSE items and the different processes by which the assessments are developed). This exercise provides evidence that supports the conclusion that quality assurance measures in place for KS2 marking (briefly described above) are effective.

### **Results and the ‘saw-tooth’ effect**

Following the marking window, in which scripts from 580,000 pupils (approximately 3.5 million papers) were marked in three weeks, KS2 test results were made available to schools on time on 4 July 2017, alongside national results and raw-score-to-scaled-score conversion charts.

Results showed that nationally, 61% of pupils reached the expected KS2 standard in all subjects (reading, mathematics and writing) compared to 53% in 2016<sup>2</sup>. 71% of pupils met the expected standard in reading and 75% in mathematics (both up five percentage points on 2016). 77% met the expected standard in grammar, punctuation and spelling (up by four points on 2016).

After a new test is introduced, we expect to see results going up in the second and third years. This improvement is normally due to pupils and teachers becoming more familiar with the content and style of the new tests. In addition, pupils taking KS2 tests in summer 2017 would have been taught an extra year of the new National Curriculum compared to pupils taking the tests in 2016. Improvements are typically rapid initially, followed by smaller changes in later years. So, it is difficult, in these early years, to disentangle the extent to which improved outcomes, in this case a 4-5 percentage point increase, could be due to better teaching and learning, or increasing test familiarity. Similar effects in relation to qualifications are known as the 'saw-tooth' effect and Ofqual's research on this can be found [here](#).

The proportion of pupils reaching the expected standard for KS2 writing increased by two percentage points, from 74% to 76%. While new writing assessments were introduced in 2016, writing is teacher assessed so the size of a saw-tooth effect may be different.<sup>3</sup> See the section below for more details on KS2 writing assessment and moderation.

KS1 outcomes are also teacher assessed, informed by externally set, but internally marked, tests. Outcomes increased from 2016 to 2017, but to a lesser degree than at KS2 (one percentage point in science, two in reading and three in mathematics and writing). Again, the size of a 'saw-tooth' effect may be different here in comparison to external tests. Outcomes of both statutory EYFSP assessments in reception and phonics assessments also slightly increased (by 1 and 1.4 percentage points respectively) in 2016. However, neither of these assessments has seen recent changes so would not have been impacted by a 'saw-tooth' effect.

### **Security of confidential assessment materials**

In 2017 there were no security breaches of confidential assessment materials, which were appropriately secured during test development and delivery. Following two breaches in 2016, STA took additional steps to secure materials, holding two security reviews, one internally focused and one undertaken by its marking supplier. Based on these reviews, it strengthened its security management and information

---

<sup>2</sup> Each pupil included in this measure met the expected standard in all three subjects. That is why this combined figure is lower than for each individual subject, as some pupils may have met the expected standard for one or two subjects, but not all three.

<sup>3</sup> Research to date on the 'saw-tooth' effect has considered external testing rather than teacher assessment.

assurance arrangements for confidential test materials across both internal and supplier processes.

Some pupils, for example due to illness or other unavoidable events, are able to take KS2 tests up to one week after the main testing window, so it is important that test content is kept secure until that extended period is complete. STA provide advice to schools and work in advance with key media outlets to advise them of this. During and immediately following test week in 2017, there was some social media activity referencing test content. STA monitored this and worked with media outlets and social media users, asking them to remove such content. Whilst incidences in 2017 were on a similar scale to 2016, they attracted a wider audience, as not all social media users removed posts. The limited volume of this activity, combined with the very small proportion pupils using timetable variations, means that it is unlikely that this activity would have impacted on test performance nationally. Nevertheless, given the increasing use of social media, STA is reviewing its internet monitoring strategy for future years and collecting additional data relating to pupils using timetable variations to enable closer oversight in future years.

### **Moderation**

In response to concerns from some stakeholders about the consistency of local authority moderation of teacher assessment, particularly for KS2 writing, we decided to carry out ongoing monitoring and research in this area. We observed KS2 writing moderation in a small number of local authorities in 2017 and interviewed those involved. This research aimed to provide insights that could be helpful in understanding more about why and in what ways there may have been inconsistencies. Our work in this area is ongoing, but we shared our initial findings with STA so that it could inform their approach for 2018. STA had already intended to make changes for 2018, and used our research to inform changes to moderation processes, moderator training and communications. We are continuing to monitor this area.

## **Section C: Content validation study and subsidiary report**

The new suite of key stage tests introduced in 2016 were based on the new primary national curriculum, introduced from 2014. In the year of their introduction, we decided to carry out a 'content validation study' to help us understand the extent to which the new KS2 reading and mathematics tests effectively sampled from the full range of learning outcomes set out in the new English and mathematics curriculum. Effective sampling is important so that pupils can gain marks for demonstrating skills, knowledge and understanding across the full curriculum and so that assessment can support effective teaching and learning; this is particularly the case for 'high-stakes' tests, such as those at KS2 which are heavily relied upon in performance measures, to mitigate the risk of teachers preparing pupils less well for aspects of the curriculum that are underrepresented within the tests.

Our study concluded that STA's approach to curriculum sampling was robust. Their approach compares favourably with approaches adopted for similar tests internationally. Given STA's interpretation of the national curriculum framework document, the Test Framework documents translate national curriculum teaching requirements into plausible blueprints for testing.

As we were designing the content validation study, we recognised that it had the potential to shed light on concerns that had been expressed about the accessibility of the 2016 reading test, so we decided to carry out a subsidiary review of evidence and data to consider this further. While the 2017 reading test did not give rise to similar accessibility concerns, our review suggested that areas relating to the test development process, including the timing of the test and review processes, could benefit from further consideration to help prevent such concerns arising in future.

STA is considering in detail the findings of both the main content validation study and the reading review of evidence. In relation to the reading review, STA has committed to further investigation into reasons why some pupils do not finish the tests and, following this, making more explicit the extent to which the speeded nature of the test is a requirement to assess reading fluency. STA will also be reviewing (a) its expert review process to ensure it provides sufficient challenge and adds the intended value; and (b) its reading text selection process to provide additional quality assurance earlier in the process.

Since our research has been published, STA has explained to us their approach to identifying potential bias in relation to reading items and how it is appropriate for key stage testing. While their approach is professionally recognised, further academic exploration of the best ways to approach the investigation of item and test bias would be helpful. As research develops, STA is committed to supporting this and remaining in line with best practice.

In relation to the points raised by the main validation study, STA intends to ensure that further work on the cognitive domains for reading and mathematics informs any future review of test frameworks. Expert research capability will need to be dedicated to this kind of developmental work to ensure that the approach to test development remains in line with international best practice.

Both studies are available [here](#) and we will continue to monitor STA's response to them.

## Section D: Regulatory framework consultation

During 2017 we reviewed and launched a consultation on our regulatory framework for National Assessments. Our regulatory framework sets out our role and responsibilities in relation to national assessments and our expectations of bodies responsible for designing, delivering and monitoring the assessments. The existing framework was published in 2011. We wanted to review it to

- provide greater transparency and clarity about how we regulate, by giving more detail about what regulation looks like in practice and by making the framework document easier to read and navigate;
- making sure our expectations of responsible bodies focus on outcomes such as validity, including by making more explicit our expectations relating to assessment purpose and strengthening our expectations about risks that should be escalated to us by responsible bodies;
- bring it up to date, in particular by reflecting changes to the bodies responsible for developing and delivering national assessments.

Our consultation will close on 20 December 2017. After considering consultation responses, we intend to publish a revised framework in spring 2018, alongside our decisions and an analysis of consultation responses. Our consultation is available [here](#).



## Section E: Future assessment arrangements

Part of Ofqual's regulatory role involves providing advice to the Department for Education in relation to proposed changes to national assessments. During 2017, the Department consulted on the future of primary assessment in two related consultations<sup>4</sup>. Key proposals included:

- Introducing a new reception baseline assessment to inform a school progress measure and, over time, removing statutory KS1 assessment in all-through primaries.
- Making changes to teacher assessments, including updating the EYFSP, removing the statutory requirement for teacher assessment of reading and mathematics at KS2, improving remaining key stage teacher assessment frameworks and in the longer term, exploring peer moderation and comparative judgement for writing.
- Making pre-key stage assessments statutory for pupils engaged in subject-specific learning but not working at the level of national curriculum assessments and introducing a new approach to assessing pupils not engaged in subject-specific learning.

We provided advice to inform government decisions related to these matters, and published our response to the consultations<sup>5</sup>.

Our response focused on the need for new assessments to be introduced with care and caution, emphasising the need for trialling, piloting and evaluation. We also emphasised the need to mitigate risks to validity in the teacher assessment model, where such assessments inform high-stakes judgements about school performance.

Decisions arising from the consultation do not significantly differ from the proposals consulted on, and have the potential to improve the validity and reliability of assessments. We welcome the intention to move to a more settled system and the recognition that improvements needed to be made in some areas. We recognise the efforts STA has made to engage key stakeholders through the consultation process and the intention to introduce changes with care and caution. In particular, we welcome the intention to carry out a full pilot and evaluation of the reception baseline, the multiplication tables check and the assessment of pupils not engaged in subject-specific learning, to make sure that the approaches are appropriate prior to roll-out.

---

<sup>4</sup> Primary Assessment in England <https://www.gov.uk/government/consultations/primary-assessment-in-england> and Rochford Review Recommendations <https://www.gov.uk/government/consultations/primary-school-pupil-assessment-rochford-review-recommendations>

<sup>5</sup> <https://www.gov.uk/government/publications/primary-assessment-consultations-ofquals-response>

Decisions made following the primary assessment consultation will mean considerable change for the primary sector over coming years and much of the detail relating to implementation has yet to be determined. It will be important that assessment validity is safeguarded and risks are carefully minimised, particularly for high-stakes teacher assessments. We welcome STA's commitment to trialling alternative approaches within the teacher assessment model, including peer-to-peer moderation and comparative judgement for writing. We would encourage further review of the teacher assessment model and the assessment of writing in particular, including consideration of alternatives and controls over both the environment in which work is produced and judgements that are made so that validity and reliability can be maximised. We welcome the decision to consider further the approach to infant, junior and middle school accountability; our view is that, if KS1 assessments were to be retained for this, assessment arrangements would need to be redesigned in order to effectively meet that purpose.

We look forward to continuing to engage with STA and relevant teams in the Department for Education as plans are developed. We will continue to monitor and seek stakeholder feedback as changes to teacher assessment, including to the EYFSP, bed-in and look forward to continuing to play our part in informing public debate on teacher assessment through our research and reporting.

## Looking forward

Our priorities for 2018 take account of a number of factors, including: the findings of our work in 2017, stakeholder views, and decisions recently made by government in relation to future changes to primary assessment. These things will inform our overall assessment of the areas of greatest risk to validity. While our focus may change in response to events or new information, our key priorities for 2018 are likely to include:

1. Monitoring changes to assessments and the introduction of new assessments, such as the development of the reception baseline and multiplication tables check.
2. A continued focus on those areas where we see risks to validity, such as teacher assessments used for high-stakes purposes.
3. Continued monitoring of those processes which can support validity, such as the psychometrically driven test-development process and standards maintenance model.

We look forward to reporting on our regulation and reflecting on national assessment development and delivery in 2018.

We wish to make our publications widely accessible. Please contact us at [publications@ofqual.gov.uk](mailto:publications@ofqual.gov.uk) if you have any specific accessibility requirements.



© Crown copyright 2017

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. To view this licence, visit <http://nationalarchives.gov.uk/doc/open-government-licence/version/3> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: [publications@ofqual.gov.uk](mailto:publications@ofqual.gov.uk).

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

This publication is available at [www.gov.uk/ofqual](http://www.gov.uk/ofqual).

Any enquiries regarding this publication should be sent to us at:

Office of Qualifications and Examinations Regulation

Spring Place  
Coventry Business Park  
Herald Avenue  
Coventry CV5 6UB

Telephone 0300 303 3344

Textphone 0300 303 3345

Helpline 0300 303 3346