

National assessments regulation

Annual report 2018

ofqual

Contents

| | |
|--|-----------|
| Executive summary | 3 |
| Introduction | 4 |
| About national assessments regulation | 4 |
| Context for 2018 | 4 |
| Section A: Priorities for 2018 | 5 |
| Section B: Monitoring assessments in 2018 | 6 |
| Test development | 6 |
| Marking quality | 7 |
| Standards maintenance | 8 |
| Results | 9 |
| Malpractice | 10 |
| Moderation of teacher assessment | 11 |
| Section C: Research | 14 |
| Update on previous research | 14 |
| Current research | 15 |
| Section D: New regulatory framework | 17 |
| Looking forward | 18 |

Executive summary

This report explains how we have regulated national assessments in 2018. We continued to take a risk-based approach, focusing on key aspects of validity and high-stakes assessments, such as those at key stage 2 underpinning school accountability. Our report describes how we are meeting commitments made in our corporate plan for 2017/18 and reports on additional work, such as our review in relation to preventing malpractice and the revision of our regulatory framework.

For national testing, this report provides assurance in a number of areas. First, we are satisfied that the Standards and Testing Agency (STA) took an appropriate approach to making sure that the new standards set in 2016 were effectively maintained for both the 2017 and 2018 tests. Second, our analysis suggests that the consistency of STA's external marking remains very high with 99.4% of markers agreeing with the definitive mark across 6.1 million marked items. Third, our observations of test development meetings suggest a strong focus within STA on ensuring the validity of each national test produced.

In relation to teacher assessment, new criteria were introduced for writing at both key stages 1 and 2, meaning that comparisons of national standards in writing between 2018 and previous years cannot be made.

During the year, we continued to monitor key areas of risk to validity and provide feedback where improvements could be made, for example in relation to the moderation of key stage 2 writing assessment and the prevention and detection of malpractice. STA has responded to our feedback and relevant actions being taken in these areas are set out in this report.

This year we have begun research on international approaches to the assessment of writing at the end of the primary phase. We hope this will usefully inform stakeholder debate and STA's current exploration of alternatives to the present approach to the assessment of writing. Our report sets out some early findings of this research, which we look forward to publishing during the first part of 2019.

During 2017/18 we consulted on and published a new regulatory framework to help explain our national assessments role. We also wrote to the Education Select Committee, to which we report, about our approach to regulation in light of proposed changes to STA's delivery model. Finally, we continued to provide technical advice on new assessments being developed, in line with our response to the government's primary assessment consultations in 2017.

The academic year 2018/19 will see changes, including revisions to teacher assessment frameworks at KS1 and the removal of some KS2 teacher assessments. Our regulation will continue to focus on monitoring areas of risk, providing technical advice and reporting on processes critical to supporting national assessment validity.

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 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). 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 2018

2018 was the third year of a new suite of assessments in mathematics and English. These 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). Revised teacher assessment frameworks in writing at both KS1 and KS2 were introduced for first use in 2018.

There was also a KS2 science sample test taken in 2018. This test is administered biennially by STA in a selection of schools to inform a national view of standards in primary science. Results are due to be released in 2019. Individual results are not provided to schools or pupils. The science sample test is next due to be administered in 2020.

Section A: Priorities for 2018

During the 2018 assessment cycle, we continued to focus on monitoring procedures that are critical to supporting valid test outcomes, including test development, standards maintenance and marking processes. We provided technical advice on the prevention and detection of malpractice and on new assessments in development such as the reception baseline and multiplication tables check, in line with our [response](#) to the primary assessment consultations (published in June 2017). We continued to focus on KS2 assessments, as the highest-stakes national assessments that form the basis of progress and attainment school accountability measures.

In 2018 we monitored and provided feedback to STA about moderation of teacher assessment, following [research](#) on the consistency of moderation of KS2 writing in 2017 and in light of the introduction of revised teacher assessment frameworks for writing for 2018. We monitored STA's response to previous research, including our moderation research and our content validation study and subsidiary [research](#) carried out in 2016-17 on KS2 testing. We also began research on international approaches to writing assessment at the end of the primary stage.

In March 2018, we published a revised [regulatory framework for national assessments](#), following our [consultation](#) about this in the autumn of 2017. Our revised framework aims to provide greater transparency and clarity about how we regulate national assessments and make sure our expectations of bodies responsible for national assessments focus on outcomes, particularly validity. In September 2018, we wrote to the Education Select Committee, to which we report, about our regulatory approach, including during an upcoming period of business change as STA transitions to a new test operation supplier for 2020.

This report summarises our activities and provides a view on key aspects of the validity of national assessments in 2018.

2018 priorities: summary

1. Monitoring processes supporting the validity of summer 2018 tests, in particular test standards maintenance and marking (Section B)
2. Publishing our research on the moderation of teacher assessment of writing at KS2, monitoring STA's response to our research and beginning a study on international approaches to the assessment of primary writing (Section C)
3. Consulting on and publishing our new national assessments regulatory framework (Section D)

Section B: Monitoring assessments in 2018

Test development

National curriculum test development is a complex technical process taking approximately 3 to 4 years. Tests are required to meet all the requirements of the relevant [Test Framework](#), including sampling appropriately from the national curriculum, providing effective differentiation across the range of pupil performance within the national cohort, providing appropriate accessibility and meeting diversity and inclusion requirements. To support this, a process of expert review, prior to constructing tests is followed 3 times over 3 years for all tests in all subjects at key stages 1 and 2:

- Prior to the first item (question) trial (Item Validation Trialling, or IVT)
- Prior to the second trial (Technical Pre-Test or TPT) a year later
- Prior to the live test a year later

The test development process, including expert review, allows for judgements about the content of live tests and mark schemes to be informed by a wide range of evidence collected over a 3-year period including: real pupil answers to questions as they are trialled; item functioning data; and the views of practising teachers, inclusion and disability experts, expert markers and test developers.

During the past year we have observed a number of test development meetings (12) at both KS1 and KS2; across a range of subjects (reading, mathematics, science and grammar punctuation and spelling); supporting the development of tests from 2018 to 2020 and beyond. We did not observe all types of meetings held, but the meetings we observed included:

- Expert review: Inclusion – where experts in different types of special educational needs and disabilities, SENCOs and cultural inclusion experts provide views on items and texts in development
- Expert review: Test Review Group – where teachers and subject experts provide advice on items/texts in development
- Resolution – where views of expert reviewers, such as those above, are reviewed and any issues arising are resolved
- Item Finalisation – where final decisions are made about items/texts to be trialled
- Test Construction – where final decisions are made about which items/texts should be included in tests to be trialled or used for live tests
- Mark Scheme Finalisation – where decisions are made on the final content of mark schemes informed by item data and expert marker judgements
- Project Board – governance meetings where senior staff are presented with validity evidence and outcomes of the test development process at key milestones, including to sign off final tests for trialling and live use

At the test development meetings we attended, we observed expert and informed discussions. STA's test developers, who are directly responsible for creating particular tests, displayed extensive knowledge of their subject; many had experience as teachers. We observed careful attention being paid to Test Framework requirements: as well as checks from test developers and reviews of

item classifications during expert review meetings, these requirements are systematised by psychometricians (specialist assessment statisticians) into algorithms to provide additional assurance that together, all the items in the test appropriately cover the curriculum.

Throughout the process, discussions and decisions were informed by real answers given by children during the pre-test process. We observed careful discussion about how pupils would respond to questions in live testing. While it was not always clear that all external reviewers added significant value, many provided relevant and useful feedback and we observed test developers asking probing questions aiming to elicit views from external reviewers. Since the set of meetings on which these observations are based, STA has reviewed its expert review process as described in Section C below.

As well as considering which questions and texts should be included in the test, we observed test developers considering carefully the appropriate order of questions, including how the ramping in difficulty of questions through the tests should work. We also observed numerous discussions about diversity, inclusion and the accessibility of the test to children with special educational needs and disabilities, such as whether the text or layout of particular questions were sufficiently clear and accessible and how questions would be perceived and engaged with by children with special educational needs and different types of disabilities.

We regularly observed discussions which suggested that the team developing tests focuses on continuous system improvements. For example, we observed discussions about how expert comments could better be recorded and about different approaches to pre-testing that were being trialled in order to provide test developers with more potential material from which to select final tests. We also observed effective oversight from senior managers at test development governance meetings, checking that processes had been followed and challenging test developers to ensure that relevant considerations had taken place.

In conclusion, our observations of these test development meetings suggested a keen focus on ensuring assessment validity in every test produced. More information on recent and planned improvements being made to test development processes are set out in Section C below.

Marking quality

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 providing results that can be relied on.

For 2018, the same measures were in place to assure marking quality as there had been in the previous year. These include:

- 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 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

In 2018, as for 2017, we observed KS2 marker training for reading, mathematics and grammar, punctuation and spelling tests (GPS). Approximately 3,400 markers were trained on Saturday 19 May (immediately after KS2 test week) at 37 different events across the country, using a common script and training materials. During the day, markers were trained on how to make judgements for each item they would be marking. Both acceptable points (creditworthy responses) and non-creditworthy responses had been identified in advance and were explained in detail, using real pupil responses from pre-testing. This aimed to provide clarity and exemplify complex points. During training, markers were asked to complete training exercises, overseen by their team leaders, to ensure they understood the approach required. Team leaders we observed were able to answer markers' questions confidently and clearly. What we observed suggested a keen focus on ensuring consistency between markers.

We also observed the selection of 'validity items' for KS2 reading, mathematics and GPS. Validity items are pupil responses that have been pre-marked by senior markers and are then distributed amongst real items during marking to check markers are marking correctly. During the 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 monitoring and assuring marking quality. As last year, items were selected to check that specific aspects of the mark scheme, marking principles and training were understood for each question in the test. We saw validity items being selected across a range of acceptable points and at each mark-point (so that markers would be tested on validity items worth each of the different marks available, for example 0, 1, 2 and 3 marks). We observed careful consideration of the selection of items that would pick up on possible marker mistakes. We also saw items being rejected that were considered too straightforward to mark to be a useful check of marking quality.

We analysed operational marking data using the same methodology as we did in 2017. Our analysis suggested that the consistency of STA's external marking remains very high, with 99.4% of markers agreeing with the definitive mark across 6.1 million marked items.

Standards maintenance

Each time a new test is produced its exact level of difficulty is likely to be slightly different to previous tests. So any test developer, such as STA, must ensure that there is a technically appropriate process for maintaining standards to ensure that the meaning of the test result remains consistent between different tests.

To maintain test standards each year, STA use a psychometric (statistical) process called 'equating'. This is supported by significant pre-testing of items used in live tests over a number of years, alongside anchor items, for which the standard is already known. The equating process is explained in our national assessments report for 2017, available [here](#).¹

¹ Technical detail on the equating process is set out in Section 13 of the STA's Test Handbook, available [here](#).

The process for maintaining test standards in 2018 was based on the same assumptions and professionally recognised techniques as in 2017. We reviewed these assumptions in 2017. We observed standards maintenance meetings for both KS1 and KS2 tests in 2018. We are satisfied that STA took an appropriate approach to making sure that the new standards set in 2016 were effectively maintained for both the 2017 and 2018 tests.

The tests in 2018 were overall of a similar level of difficulty to 2017 tests. Some were marginally less difficult than in the previous year, for example, KS2 mathematics and grammar, punctuation and spelling, and KS1 reading were slightly easier than they had been in 2017. 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 because of the standards maintenance approach taken. This means that results each year reflect the attainment of pupils, rather than the level of difficulty of the particular test that was taken.

Results

Following the marking window, in which scripts from more than 600,000 pupils (approximately 3.5 million papers) were marked in 3 weeks, KS2 test results were made available to schools on 10 July 2018, alongside national results and raw-score-to-scaled-score conversion charts.

In KS2 tests, 76% of pupils met the expected standard in mathematics and 78% met the expected standard in grammar, punctuation and spelling (GPS). Both of these increased 1 percentage point on the previous year. 75% of pupils met the expected standard in reading, a 4 percentage point increase on 2017 outcomes.

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 and is known as the ‘saw-tooth’ effect (Ofqual’s research on this in relation to qualifications can be found [here](#)). In addition, pupils taking KS2 tests in summer 2018 would have been taught an extra year of the new National Curriculum (4 years in total) compared to pupils taking the tests in 2017. ‘Saw-tooth’ improvements are typically rapid initially, followed by smaller changes in later years. It is difficult, in the first few years of a new test, to disentangle the extent to which improved outcomes could be due to better teaching and learning or increasing test familiarity. Similarly, it is not easy to determine the reasons for the difference in the amount of any ‘saw-tooth’ increase in different test subjects.

Turning to teacher assessment, revised assessment frameworks for writing were introduced in summer 2018. Significant changes were made to assessment criteria, including the introduction of a “more flexible” approach to the ‘secure-fit’ framework. The revised frameworks removed some of the previous criteria and added other elements; some criteria also changed significantly in their focus. Changes, no matter how small, to ‘secure-fit’ or ‘mastery’ assessment criteria will make at least some change to the overall assessment standard. For instance, because of the additions, removals and changes, it is possible that some pupils who did not meet the 2017 standard for KS1 or KS2 writing would have met the 2018 standard, and that some pupils who met the 2017 standard would not have met the 2018 standard.

While the impacts on pass rates for changes of this nature cannot easily be predicted, results may change overall (go up or down) or remain broadly stable. For this reason, pass rates cannot be directly compared between old and new frameworks. The proportion of pupils reaching the expected standard for KS2 writing in 2018 was 76%. While in 2017 this percentage was 74%, it is not possible to draw conclusions about changes in pupil attainment in writing at KS1 or KS2 between 2017 and 2018, owing to the change to assessment criteria, and therefore to the standard.

This caveat also applies to the combined figure reported for reading, maths and writing attainment²; which in 2018 showed that 64% of pupils reached the expected KS2 standard in all subjects (reading, mathematics and writing). In 2017 the combined figure was 61%. However, this change does not necessarily demonstrate that underlying pupil ability has improved such that this would be generalisable to other testing situations. Indeed, due to changes to the writing teacher assessments and the likely continuation of some 'saw-tooth' effect in testing, it is not possible to draw straightforward conclusions as to whether performance standards are rising based on 2018 assessment data.

KS1 outcomes are teacher assessed, informed by externally set, but internally marked, tests in reading and mathematics. Having seen small increases from 2016 (when new assessments were introduced) to 2017, outcomes in 2018 remained broadly as 2017 with 75% meeting the expected standard for reading, 76% for maths and 83% for science. Again, for writing, new teacher assessment frameworks were used for the first time in 2018, so results (70% meeting the expected standard) are not comparable with 2017. Outcomes of 2018 phonics assessments also slightly increased, by 1 percentage point from 2017. The phonics assessment has been in place since 2012 and after an initial period of increase, which may have been partly related to increasing test familiarity, outcomes have remained broadly similar since 2016. A similar pattern can be found in relation to outcomes of the Early Years Foundation Stage Profile (EYFSP) which is an observational teacher assessment at the end of the reception year (when children are aged 4-5). After an initial period of increase following the introduction of a revised Profile in 2013, the rate of increase has slowed in recent years. This year 71.5% of children achieved the 'good level of development' standard, up 0.8 percentage points on 2017.

Malpractice

STA have well-established processes in place for investigating reports of maladministration³. Investigations can lead to the annulment of results for a pupil, group of pupils or whole cohort where the accuracy or correctness of test results is in doubt. Resultant disciplinary action may be taken by other bodies, for example, by school governing boards and/or the Teacher Regulation Agency. Investigations into

² 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.

³ STA investigates reports of malpractice as well as maladministration but does not determine whether acts were accidental or intentional. Malpractice is used in this report to mean intentional maladministration. STA defines 'maladministration' as "any act that could jeopardise the integrity, security or confidentiality of the national curriculum assessments and could lead to results that do not reflect the unaided abilities and achievements of pupils." – this definition includes both accidental maladministration and intentional malpractice/cheating.

<https://www.gov.uk/government/collections/maladministration-reports>

malpractice and maladministration are vital to maintaining test validity so the public can have confidence in test results and to deter future malpractice.

While the number of malpractice and maladministration complaints made to STA represents only a very small fraction of assessments taken, reports of malpractice can have a negative impact on confidence in the assessment system. This summer saw a number of media reports relating to malpractice in key stage testing. Although published maladministration data had not suggested an increase, in light of the potential impacts on public confidence, during spring and summer 2018 we reviewed available documentation relating to STA's approach to malpractice prevention and detection and provided feedback on areas which had the potential to be strengthened. This included in relation to test administration, teacher assessment, safeguarding of confidential assessment materials and the use of special considerations and access arrangements.

STA have considered our feedback and are putting in place a number of measures, some of which they had already intended to implement prior to our review. These include:

- Better identification and monitoring of potential conflicts of interest, for example in relation to experts with access to confidential assessment material
- Additional data collection and analysis to identify trends and anomalies to enable follow-up where appropriate, for example:
 - improving recording so that trends can be better monitored
 - identifying schools with consistently unusual patterns or anomalies in test, teacher assessment or phonics check outcomes data
 - monitoring unusual or disproportionate use of access arrangements
- Strengthening guidance and communication to support deterrence, for example, adding requirements around conflicts of interest management to local authority monitoring guidance

We also suggested that STA may wish to consider strengthening their current recommendation, that schools should arrange for KS2 tests to be independently observed, into more of an expectation or requirement to support verification of the integrity of test administration. STA are currently considering the language of their recommendation in relation to test observers.

In relation to test administration, STA were also able to tell us that local authorities monitor more than the required 10% of schools each year to ensure that proper administration processes are followed.⁴ These additional visits are carried out partly in response to intelligence and visits can themselves provide information that may lead to further investigation. STA are now also working with local authorities to moderate specific schools in response to intelligence in relation to teacher assessment processes.

Moderation of teacher assessment

Following our research into the consistency of local authority moderation of KS2 writing assessments in 2017, we recommended that STA take steps to reduce risks of inconsistency for future years, reduce the likelihood that schools can predict when they are going to be moderated and keep the approach to the assessment of writing

⁴ 2479 schools in 2018, 932 more than the 10% required

under review. STA made a number of changes for 2018 designed to improve consistency, based on both their own evidence gathering and our analysis. The main changes planned for 2018 were:

- Introducing new teacher assessment frameworks for writing
- Making improvements to moderator training and the moderator standardisation test
- Improving the speed and responsiveness of communications to teachers and moderators
- Encouraging local authorities to moderate more than the minimum 25% of schools required in order to reduce the predictability of moderation
- Considering whether additional feedback could be given to moderators by STA's external moderators

In relation to these commitments, STA broadly carried these out as intended for 2018 assessments, for example, publishing new teacher assessment frameworks, making changes to the moderator standardisation test (e.g. improving authenticity by allowing moderators to judge using paper samples of pupil writing, rather than online samples) and encouraging local authorities to moderate more than the minimum sample of schools. More materials were disseminated to support moderators' and teachers' understanding of the revised frameworks (for example, a webinar/video was produced and STA's training materials were made available to teachers). While it was not always clear that all questions asked by moderators were effectively answered, for example at training sessions in 2017, STA reorganised its helpline to provide more dedicated resource to support queries relating to teacher assessment and provided greater feedback to moderators from its own external moderation. Finally, STA also reaffirmed its commitment to explore alternative approaches to the current model, such as comparative judgement; this work is ongoing.

Following our 2017 research, we continued to monitor KS2 writing moderation for 2018. In December 2017, we observed a number (3 of 5) STA moderator training events, designed to prepare local authorities to moderate in summer 2018 using the revised KS2 writing framework. We also observed a moderation session in small number of local authorities (5) in June 2018. These observations were not carried out as part of a research project, but formed part of our ongoing regulatory monitoring.

Our monitoring of both moderator training and moderation sessions suggested that the new frameworks were generally welcomed by those teachers and moderators we spoke to, for example because they better aligned with views of what constituted effective writing. Our monitoring also suggested that risks of inconsistency in the interpretation of the new frameworks remain and that further clarity would be beneficial in future years, in particular:

- Clarifying the meaning of the advice that the new framework provides "greater flexibility", so that a consistent approach is applied in interpreting what is still designed to be a 'secure-fit' framework
- Providing a greater level of guidance in relation to the 'greater depth' standard, focusing on supporting consistent decisions at the borderline of the expected standard and the 'greater depth' standard

Based on STA's own evidence-gathering and our feedback, for 2019 STA intend to further develop moderator training, for example, by allowing more opportunity for in-

depth exploration of the standards, focusing on areas highlighted in feedback from 2018 moderation and improving quality assurance of trainers. We will continue to monitor this area.

We observed some interesting different approaches to moderation in 2018, for example, one local authority we visited used a 'warehouse' model to moderation – where schools came together in a central location and moderators worked under the supervision of a senior moderator who provide advice and support consistency. While it is not necessarily the case that this type of approach would be the most appropriate, our observations suggested that it may be worth researching different approaches taken by local authorities in order to identify whether there may be any processes or controls which have the potential to better support consistency and enhance the current approach.

Section C: Research

Update on previous research

During 2016 and 2017 we carried out a number of research studies into KS2 assessment:

- A [content validation study](#), looking at how the new primary national curriculum was translated into testing
- A [review of evidence](#) on the accessibility of the 2016 reading test
- An [observational study](#) of local authority moderation of writing

Our 2017 annual report set out STA's response to these three pieces of research and we committed to continue to monitor these areas. In relation to the third study above (local authority moderation), our ongoing monitoring and STA's response to this is described in Section B above.

In relation to the first two research studies on key stage 2 testing the STA has, since our last report in 2017:

- Reviewed its expert review processes (used to review materials in development during the 3 to 4 year test development process) and from this autumn has made changes to expert reviewer recruitment, review processes and guidelines to improve the quality of external scrutiny of test development materials and increase effective challenge. Expert reviewers already include serving teachers, subject experts and disability experts and improvements aim to increase the diversity and level of expertise of expert reviewers, eg by:
 - seeking input from people with a wider range of professional expertise in specific types of disabilities
 - targeting schools with specific expertise in SEND and diverse intakes to provide expertise to enhance existing teacher reviews
 - asking reviewers more targeted questions about the materials (for example via questionnaires) to provide additional challenge
- Reviewed its arrangements for monitoring DIF (differential item functioning⁵) with an external technical advisory panel who have confirmed that the methodology is appropriate. They have also provided additional training on fairness in assessment to test developers
- Introduced new processes to review and approve texts before questions are written related to them, in order to improve the choice and quality of reading texts. STA are also trialling new approaches to reading pre-testing to improve flexibility in final test construction and are looking at how to source a wider range of texts for future years, including more modern writing

STA are currently investigating further the numbers of pupils who omitted certain questions and did not reach the end of the reading test. They also intend to review

⁵ DIF is a type of data analysis which helps understand whether questions may be performing differently with different groups of pupils, such as girls and boys.

the cognitive domains for the various subjects as part of future test framework review work. We will continue to monitor and report on these areas.

Current research

Following our research on KS2 writing moderation in 2017, we are currently carrying out research on international approaches to the assessment of writing at the end of primary education. Our aim is to provide evidence that can both support stakeholder debate and inform government's ongoing exploration of potential alternatives to the current model. The research focuses on assessments around the age of 10/11 and covers jurisdictions which are English speaking or use English for assessments, including both state and private providers of large-scale assessments. It also provides a summary of the different approaches taken to the assessment of writing in England since national testing was first introduced at KS2 in 1995.

To date we have identified 15 large-scale assessments, which systematically assess writing around the end of the primary stage.

- Australia – National Assessment Program: Literacy and Numeracy (NAPLAN)
- Canada (Ontario) – Assessment of Reading, Writing and Mathematics: Junior Division (also known as the Junior Division Assessment; JDA)
- Caribbean – Caribbean Primary Exit Assessment (CPEA)
- England – National curriculum assessments: Key Stage 2 (KS2)
- Hong Kong – Territory-wide System Assessment (TSA)
- New Zealand – Assessment Tools for Teaching and Learning (e-asTTle)
- Pakistan – National Achievement Test (NAT)
- Philippines – National Achievement Test (NAT)
- Scotland – Scotland National Standardised Assessments (SNSA)
- Singapore – Primary School Leaving Examination (PSLE)
- Trinidad and Tobago – Secondary Entrance Assessment (SEA)
- Uganda – Primary Leaving Examinations (PLE)
- USA (California) – California Assessment of Student Performance and Progress (CAASPP)
- USA (California) – English Language Proficiency Assessments for California (ELPAC)
- USA – National Assessment of Educational Progress (NAEP)

A range of approaches to the assessment of writing are taken in these different systems. Just 2 of the identified assessments (KS2 in England and the CPEA in the Caribbean) use a portfolio of writing assessed by teachers. The majority (13 out of 15) assess writing via a test. These use a variety of assessment modes, from online adaptive to paper-based. They also use different task types, such as extended-response, short response and multiple choice or a mixture of these.

Different approaches are also taken to the skills required to be displayed. For example, some assessments require pupils to write for a specified number of different purposes, by producing a narrative, a persuasive piece and informative writing. However, some focus more on writing skills with an expectation that these can be displayed across a range of contexts. Others focus on the technical skills of

writing (such as spelling, grammar and punctuation). 9 of the 15 assessments can be described as 'high-stakes' for pupils, teachers/schools or both. Within these 9, 7 use testing, with only England and the Caribbean using teacher assessment⁶.

Our research will aim to describe and discuss the different approaches taken in international assessments, which we hope will provide useful evidence to inform current debates on the challenging question of how best to assess writing at the end of the primary stage. We will aim to publish this research in early 2019.

⁶ Although writing assessments in the Caribbean contribute to school performance, they constitute less than 7% of the measure, thus the 'stakes' of these assessments may not be as high as might otherwise be assumed.

Section D: New regulatory framework

Ofqual is required to publish a regulatory framework for national assessments. Our framework explains how we fulfil our regulatory role for national assessments, which is different to the role we have in relation to qualifications. It also sets out our expectations of those bodies responsible for national assessments. In autumn 2017 we consulted on updating our regulatory framework for national assessments; our previous framework had been published in 2011 and was due for review. While we did not propose to significantly change our approach, our consultation asked for views on our intentions to:

- Bring our framework up-to-date and reflect changes to the bodies responsible for developing and delivering national assessments
- Provide greater transparency and clarity about how we regulate, including by explaining in more detail our regulatory approach, our regulatory tools and our focus on the validity of national assessments
- Make sure our expectations of responsible bodies focus on outcomes such as validity, rather than prescribing certain administrative approaches
- Make more explicit our expectations relating to assessment purpose and to strengthen our expectations about risks that should be escalated to us by responsible bodies

We received 12 responses to our consultation from a range of representative bodies, such as teacher associations, and assessment organisations. Respondents generally supported our proposed framework, although there were some specific comments suggesting our approach in some areas could be clarified. Based on these responses, we decided to implement the framework on which we consulted, with a small number of wording changes to make our approach and focus clearer in some areas. Our new framework was published and came into effect in March 2018. It is available [here](#).

Since the publication of our new framework, we have written to the Education Select Committee, to which we report, about our regulatory approach, including during an upcoming period of business transition as STA moves to a new test operations supplier from the 2020 assessment cycle. During this period of transition we will continue to focus our regulation on assessment validity, but will broaden the scope of our work to monitor for any risks to validity that may be posed by this change. This does not represent a significant change in our strategic approach and is in line with our existing regulatory framework. The DfE has confirmed to us its ongoing oversight and responsibility for test operations delivery and that transition operations are being quality assured by cross-government bodies including the Government Internal Audit Agency. Our exchange of letters with the Secretary of State and our letter to the Education Select Committee about this can be found [here](#).

Looking forward

During 2017/18, STA has continued to research and develop assessments, in particular the reception baseline assessment, due to be introduced in autumn 2020, and the multiplication tables check, due to be introduced in the summer of 2020 for pupils in year 4. There has also been development work within the wider Department for Education on the Early Years Foundation Stage Profile, with a revised Profile due to be used statutorily from 2021/22. During this period, we have provided technical advice relating to the validity of these assessments, for example, relating to the need for accessibility of the multiplication tables check, which will be the first computer-based national assessment, the need to ensure the reception baseline assessment is as robust as possible given pressures resulting from its use in school accountability measures and the need to ensure clarity in EYFSP assessment criteria to promote consistency of interpretation by teachers. More information about our views on validity in relation to these assessments can be found in our response to the primary assessment consultations, published [here](#).

For the academic year 2018/19, revised teacher assessment frameworks will be introduced for KS1 reading, mathematics and science and for KS2 science. As with the writing frameworks this year, these changes will mean that results in these subjects in summer 2019 will not be comparable with results from summer 2018. Also from 2019, teacher assessment frameworks for reading and mathematics at key stage 2 are being withdrawn from use. No such changes are planned for testing; KS1 and 2 tests in reading, mathematics and GPS remain in place.

There will also be changes to the assessment of pupils working below the level of the national curriculum tests: pre-key stage standards are being introduced in 2018/19 to replace the current P scale assessments above P scale 4. P scales 1-4 remain in place for 2018/19, but will be replaced for 2019/20. As above, outcomes of new assessments will not be comparable with those of current assessments.

Our regulation during the 2019 assessment cycle will continue to take account of the findings of our work in 2017 and 2018, stakeholder views, and changes to primary assessment arrangements. These 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 2019 and beyond are likely to include:

- Monitoring changes to assessments, the development of new assessments and potential validity impacts of operational changes
- Publication of research on international approaches to the assessment of primary writing
- A continued focus on areas where we see risks to validity, such as teacher assessments used for high-stakes purposes
- Continued monitoring of those processes which can support validity, such as the psychometrically driven test-development process, standards maintenance model and marking processes

We look forward to reporting on our regulation and reflecting on national assessment development and delivery in 2019.

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



© Crown Copyright 2018

This publication is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated.

To view this license, visit

www.nationalarchives.gov.uk/doc/open-government-licence/

or write to

Information Policy Team, The National Archives, Kew, London TW9 4DU

Published by:

ofqual

Earlsdon Park
53-55 Butts Road
Coventry
CV1 3BH

0300 303 3344
public.enquiries@ofqual.gov.uk
www.gov.uk/ofqual