

[Home](#) ▾ [Education, training and skills](#) ▾ [School curriculum](#) ▾ [Early years curriculum](#)
▾ [Reception baseline assessment development handbook](#)



[Standards
& Testing
Agency](#)

Guidance

Reception baseline assessment: development handbook

Published 7 September 2026

Contents

- [1 Introduction](#)
- [2 The assessment](#)
- [3 The assessment development process](#)
- [4 Item origination](#)
- [5 Item design](#)
- [6 Question and item review](#)
- [7 Trialling](#)
- [8 Analysis of trialling data](#)
- [9 Assessment construction](#)
- [10 Governance](#)
- [11 Assessment administration](#)
- [12 Assessment validity](#)
- [13 Glossary](#)

1 Introduction

The reception baseline assessment (RBA) became statutory in 2021 and was delivered by the National Foundation for Educational Research (NFER). In 2025, this original version was replaced with a digital version developed by the Standards and Testing Agency (STA).

STA is an executive agency of the Department for Education (DfE). STA's national assessment functions are regulated by the Office of Qualifications and Examinations Regulation (Ofqual). This is in line with Ofqual's [regulatory framework for national assessments](#).

Following the Standards for Educational and Psychological Assessment [\[footnote 1\]](#), this handbook describes how the RBA was developed for the 2025 to 2027 assessment cycles and presents validity and reliability evidence.

1.1 Purpose of this document

This handbook has been produced by STA to explain how the revised RBA has been designed, developed and delivered.

1.2 Who is this document for?

This guidance is primarily for a technical audience. It is published for reasons of openness and transparency. It will be of interest to those who develop assessments or are involved in assessment, including in schools.

2 The assessment

The RBA is an on-entry assessment which forms part of the statutory assessment arrangements for pupils in the reception year. It is distinct from the early years foundation stage (EYFS) profile which summarises and describes pupils' attainment at the end of the reception year.

The RBA is an age-appropriate assessment administered to pupils within the first 6 weeks of starting reception. It is made up of 2 components assessing:

- early mathematics
- early literacy, communication and language

2.1 Purposes and use of the assessment

The purpose of the RBA is to create a baseline measure from which cohort-level progress to the end of key stage 2 (KS2) can be calculated. The RBA is not intended to:

- be used in any way to measure performance in the early years, evaluate or compare pre-school settings or hold early years practitioners to account
- provide detailed diagnostic information about pupils' areas for development
- provide ongoing formative information for practitioners or replace any on-entry baseline work schools already undertake

The assessment falls under Ofqual's [regulatory framework for national assessments](#).

2.2 The assessment framework

The [RBA framework](#) sets out the content and cognitive domain for the RBA.

The purpose of the framework is to guide the development of the assessment. By providing consistent parameters for the development and construction of the assessment, the assessment framework allows valid and reliable assessments to be constructed when required.

The framework was written primarily for those who write assessment materials. It has been made available to a wider audience for reasons of openness and transparency. STA developed the framework in consultation with panels of teachers and subject experts, to refine the framework's content, confirm its validity and make it fit for purpose.

The framework provides information about what the assessment covers, but it does

not provide information on how teachers should teach or assess pupils' progress.

The assessment framework contains:

- a content domain, setting out what can be assessed through the assessment
- a cognitive domain, outlining the demands of the assessment and the cognitive skills required
- an assessment specification, which:
 - gives details of assessment format, item types, response types, marking and the balance of marks across the content and cognitive domains
 - explains how the assessment outcomes will be used

2.2.1 Development of the cognitive domain

The cognitive domains make explicit the thinking skills and intellectual processes required for each assessment. Each item (the smallest unit of a question for which a mark is awarded) is rated against the relevant components of the cognitive domains. By taking this information into account during assessment construction, the demand and difficulty of the assessment will be appropriate.

The cognitive domains were initially developed through a literature review. The resulting model allows items to be rated across different areas of cognitive demand.

To validate the cognitive domains, panels of teachers reviewed the assessment frameworks. Pairs of test development researchers (TDRs) also independently classified items against the cognitive domain, and their classifications were compared.

Refinements were made to the cognitive domains based on the inter-rater consistency between TDRs. This ensured the cognitive domains published in the assessment framework were valid and usable.

2.3 Delivery of the assessment

RBA is delivered one-to-one using 2 digital devices. The practitioner uses one device to navigate through the assessment, reading questions and instructions to the pupil. Depending on the question type, the pupil responds by interacting with a touchscreen device, giving their answer verbally or interacting with physical manipulatives. Responses are either automatically marked or practitioners input a response of 'correct', 'incorrect' or 'no response'.

2.4 Routing

Routing is used to prevent pupils being presented with too many items with which they are unlikely to be successful. Routing means that a more difficult question may not be presented to a pupil if their previous answers suggest they are unlikely to get it right. It also helps to reduce the time required for the assessment and could avoid the possible discomfort a pupil may feel if they are unable to answer a question. Items that the pupil answers incorrectly and items that are not presented to the pupil due to routing will score 0.

Within each component, there are several places where routing rules may be applied based on how a pupil responds to each question they are presented with. Potential routing options were considered by TDRs and psychometricians as part of the design of the technical pre-test (TPT) trial in 2022 to enable the post-trial analysis of the impact of applying routing rules to particular items or groups of items. This analysis was considered alongside the routing rules in place in the RBA administered by NFER, the age of the pupils and the purpose of the assessment.

2.5 Reporting

No pupil or school-level RBA scores are shared with schools and no national, regional or school-level data is published for RBA by STA or DfE. Instead, the raw scores are stored in the National Pupil Database (NPD) and used to create a cohort-level progress measure for schools at the end of KS2.

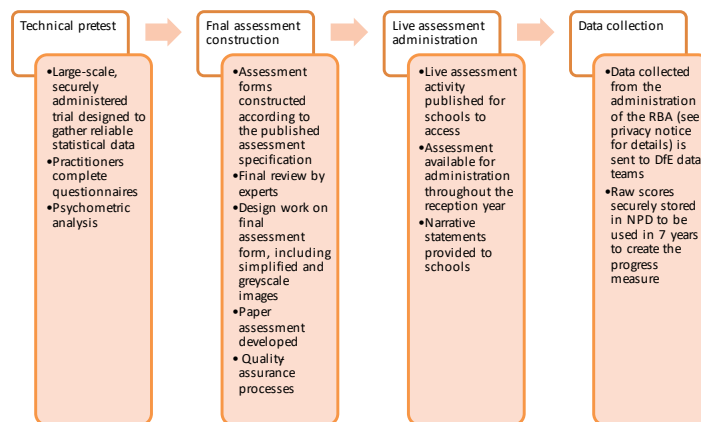
At the end of the assessment, a series of narrative statements is generated and provided to schools to describe how each pupil performed in the different content domains presented in the assessment. There is no legal requirement for schools to report RBA narrative statements to parents. However, schools must share a pupil's RBA narrative statements with the child's parents if they request them. No individual scores are reported to parents.

3 The assessment development process

The assessment development process is based on an item-banking model. Each item is taken through a series of phases to establish whether it is valid, reliable and meets the purposes of the assessment. Items meeting these criteria are considered for possible inclusion in a live assessment. Any suitable items not used in a live assessment are retained for potential use in future live assessments. Live assessments must meet the assessment specification in the assessment framework.

The assessment development process takes between 2 and 3 years to complete. An overview is provided in Figure 1 . Full details of each stage are provided in sections 4 to 13. Not all items will go through the entire process, depending on the predictability of item performance.

Figure 1: Overview of assessment development processes



Transcript of Figure 1: Overview of assessment development processes

Stage 1 - Item origination:

- Items written by test development researchers
- Items reviewed by EYFS experts

Stage 2 - Expert review 1:

- Panels of experts review items
- Amendments made in response to their feedback
- Items grouped into trial assessment forms

Stage 3 - Item validation trial:

- Securely administered trial
- Approximately 300 pupils trial each item
- Practitioners complete questionnaires
- Psychometric analysis
- Items are amended in response to outcomes

Stage 4 - Expert review 2:

- Panels of experts review items again
- Amendments made in response to their feedback
- Items grouped into trial assessment forms

Stage 5 - Technical pre-test:

- Large-scale, securely administered trial designed to gather reliable statistical data
- Practitioners complete questionnaires
- Psychometric analysis

Stage 6 - Final assessment construction:

- Assessment forms constructed according to the published assessment specification
- Final review by experts
- Design work on final assessment form, including simplified and greyscale images
- Paper assessment developed
- Quality-assurance processes

Stage 7 - Live assessment administration:

- Live assessment activity published for schools to access
- Assessment available for administration throughout the reception year
- Narrative statements provided to schools

Stage 8 - Data collection

- Data collected from the administration of the RBA (see [privacy notice](#) for details) is sent to DfE data teams
- Raw scores securely stored in NPD to be used in 7 years to create the progress measure

3.1 STA assessment development staff

Assessment development is conducted by STA's in-house Assessment Research and Development division. This team comprises assessment experts, psychometricians, TDRs, assessment designers and project staff. STA's technical specialists have a range of valid experience. This includes experience of working on a number of assessment programmes, expertise in the theories and techniques of assessment, and classroom experience.

3.2 RBA assessment services

The RBA is developed and delivered on STA's online platform, which can be accessed using [DfE Sign-in](#). This comprises 3 digital services, each designed for specific activities:

Service	Activity
Assessment service: manage your school's assessments	Schools can register pupils, record reasons for pupils not taking the assessment, complete the headteacher's declaration form (HDF), and view and download narrative statements.
Assessment service: start an	Practitioners can complete training, familiarise themselves with assessment questions, check their devices can connect,

assessment	manage accessibility settings and administer the assessment.
Assessment service: develop an assessment	STA staff can create, develop and manage RBA assessment materials. This can only be accessed by approved STA users.

3.3 RBA's item bank

Items, questions, forms and form groups are created and stored in the assessment service. Metadata and evidence associated with these entities are recorded and stored separately in STA's secure content management system. The use of an item-banking assessment development model contributes significantly to the efficiency of assessment construction and the validity and reliability of the assessments, by:

- enabling STA to meet the assessment specification criteria precisely, because assessments are constructed using information from trialling
- ensuring items that do not perform well in trialling are not included in the live assessments
- allowing selection of items in any combination to create the optimum assessment that meets the specification

4 Item origination

Item origination is the initial development process for items. This first stage of development for the RBA is undertaken in-house. Items are created within 'Assessment service: develop an assessment' (see [section 3.2](#)).

4.1 Planning for item writing

STA conducts regular reviews of all materials to ensure enough items are available. This takes into account all aspects of the assessment specification, such as coverage of the cognitive domain.

4.2 Internal item writing

Items for the RBA are drafted by TDRs. Depending on the nature of the item-writing project, they may do this with reference to specific published sources, international assessments or research evidence. Sometimes a familiar item format is used, with changes made to the target word or context. Indicative mark schemes, where required, are developed alongside the items. Once drafted, the items and mark schemes are subject to several stages of internal review. Some items will be considered suitable without changes, for some amendments may be proposed. Some items are removed at this stage, which may necessitate writing additional new items.

Following amendments to items, a draft is circulated to EYFS curriculum advisors who provide comments via a written report.

4.3 Mark schemes and guidance

The majority of items in the RBA only have one correct answer and, due to the nature of the constructs being assessed, these answers are unambiguous. Where different responses are permissible, marking guidance has been developed throughout the item development process to ensure that practitioners are able to easily decide whether the response is correct or incorrect.

Pupil responses for some items are collected during trialling and reviewed by TDRs. A selection of these, both correct and incorrect, are added to the training materials provided to schools. TDRs discuss creditworthy and non-creditworthy responses with stakeholders at the expert review panels.

4.4 Item classification

During the item origination process, items are created and securely stored within STA's 'Assessment service: develop an assessment.' Here they are assigned a unique identifier code. Additional item-level metadata is stored securely outside of the assessment service (see [section 3.2](#)). This is used to classify the items and monitor the characteristics of selected items against the assessment specification.

Item metadata may change as the item moves through the development process. Further details of the classifications used can be found in the [RBA framework](#).

5 Item design

Item design refers to an item's appearance and the conventions used to maintain design consistency within, and between, items.

5.1 Interfaces and processes

Design begins during item origination (see [section 4](#)). It is an important consideration throughout the assessment development process because of its contribution to pupil experience, item accessibility and item validity, including the avoidance of construct irrelevant variance (see [section 6.3.3](#)).

Item templates have been created in the assessment service for each item type in the RBA. TDRs use these to create items for the assessment to ensure consistency across items.

Following each expert review (see [section 6](#)), STA may amend item wording, layout or graphical elements. TDRs carry out amendments to the items, and STA's design team commissions additional images from external illustrators as necessary. Once amended, items are placed in forms for use during in-school trialling and the live assessment.

5.2 Key considerations

Item design can affect validity. Items that are poorly laid out are difficult to comprehend and access.

All STA assessment materials are developed to conform to relevant design principles, such as:

- ensuring the purpose of each task is clear

- ensuring tasks address assessment criteria
- avoiding a requirement for pupils to demonstrate skills that are not essential to the subject being assessed

Stylistic conventions have been developed to ensure consistency within and between the items of a particular subject, and also between subjects as far as is practicable.

Consistent use of writing style and design guidelines are also important for the item-bank development process (see [section 3](#)).

5.3 Style and design guides

Style and design guidelines provide a basis for item development so that items from different item-writing agencies are consistent. The guidelines are active documents, updated to reflect changes in the use of language and design resulting from feedback from the trialling and development process.

5.4 Copyright

All illustrations in the RBA are directly commissioned for the purpose of this assessment, so the copyright is transferred to STA.

6 Question and item review

Item review is the process of checking the content, wording and layout of questions and items to ensure appropriateness, validity, clarity and accuracy.

6.1 Stages of review

Items and their mark schemes are reviewed throughout the assessment development process by a wide team of internal and external parties.

The expert review process involves independent quality assurance of assessment materials. It is critical to the validity evidence that supports the assessments.

Reviews usually take place:

- before the item validation trial (IVT) – expert review 1
- before the technical pre-test (TPT) – expert review 2
- after the live assessment is constructed – expert review 3

The expert panels review and comment on the suitability of assessment items, identify possible issues and suggest improvements.

Security is of critical importance during the expert review process, and all those involved in the process are required to sign confidentiality agreements.

6.1.1 Panel definition and set-up

Three separate expert panels review the items at both expert review 1 and expert review 2. Each panel is provided with guidance about the purpose of their review and the intended outcomes of the meeting. Panels are shown how the items will appear on screen for pupils and the digital interactions required.

These panels are recruited through recruitment emails to schools, local authorities, subject associations and universities. The panels are regularly updated with new reviewers.

The expert review panels are chaired by a TDR.

6.1.2 Teacher panel

Teacher panels are designed to gather feedback on how materials reflect current classroom practice. The panels comprise practising teachers, headteachers, or teachers who have recently left the profession. They have experience of teaching the target age group for the assessment, meaning that STA is gathering relevant and valid advice. The panels consist of individuals of differing experience, from differing school types and geographical locations so that a range of views is gathered. The panels represent, as far as is possible within a small group, the range of teachers and school types that will encounter the live assessments. The panel is asked to comment on:

- wording of items
- question contexts, including names
- question layout and design
- digital interactions

- amount of processing involved per mark
- whether the materials are of appropriate difficulty for the age group, ability range and purpose of the assessment

The teacher panel receives the materials for review in secure conditions on the morning of the meeting. Members are asked to comment on the suitability of draft materials and to suggest any improvements that would make it a fairer assessment. After reading the materials, the panel works as a group to comment on each individual item.

6.1.3 Assessment review group

The assessment review group (ARG) comprises of assessment experts which may include:

- subject specialists
- local authority advisors
- multi-academy trust advisors
- practising teachers and headteachers who have previously served on teacher panels

Some of the ARG panel members are replaced for each assessment cycle, to ensure panels are refreshed and different people are seeing the materials, but that some consistency of review remains.

The group comments on:

- the technical accuracy of the content
- whether the materials are of appropriate difficulty for the age group, ability range and purpose of the assessment
- whether the materials reflect relevant aspects of the assessment framework

Both ARG and inclusion panel members (see [section 6.1.4](#)) receive assessment materials and a report template at least one week before the meeting takes place. Inclusion reviews may be done by report only. Panellists review each item, identify any issues and consider improvements in advance of the meeting. At the meeting, they provide feedback on each item, listen to each other's perspectives and attempt to reach a consensus on each item.

6.1.4 Inclusion panel

The inclusion panel is composed of experts in teaching pupils with special

educational needs and disabilities (SEND) and English as an additional language (EAL). Typically, the panel includes at least one each of:

- an expert in teaching pupils with visual impairments
- an expert in teaching pupils with hearing impairments
- an expert in communication and language
- an educational psychologist
- those with expertise in teaching pupils with autistic spectrum disorders
- those with expertise in teaching pupils with EAL
- an equity advisor

The panel reviews the accessibility of the standard assessment for pupils with this range of needs.

6.1.5 Decision making after external review

Resolution meetings are held shortly after expert review meetings 1 and 2.

Resolution meetings are chaired by the test developer leading that particular assessment, and are attended by:

- other TDRs or research assistants for the subject
- the deputy director, head of assessment development research or head of assessment research and psychometrics
- a psychometrician or lead psychometrician
- a project co-ordinator
- an assessment researcher (optional)

The meetings consider all of the evidence from the expert review panels and resolve all the issues raised through the expert review stage, deciding what changes will be made to the materials as a result. Evidence from previous trialling may also be considered if it is available.

The changes suggested by different reviewers may contradict each other. The lead TDR is responsible for deciding which recommendations to implement, taking all the evidence into account. If no solution can be found, the item is likely to be removed from the assessment development process.

6.1.6 External review at the end of the development process

A single expert panel reviews the items at the end of the assessment development

process (expert review 3). It comprises members of the teacher panel and assessment review group. Expert review 3 is to review the suitability of the constructed assessment, with specific consideration for overall difficulty and breadth of coverage of the content domain.

At expert review 3, the panel receives the assessment forms and guidance documents on the morning of the meeting. The panel is given time to read and work through the assessment forms before providing whole-group feedback on the suitability of the assessment. Assessment materials are not sent out in advance of expert review 3. This is to ensure the live assessment remains secure and to allow panellists to review the materials as they would on assessment day.

6.2 EYFS advisors

EYFS advisors are recruited to support the quality assurance of items throughout the assessment development process. They work alongside STA's TDRs to ensure the assessments provide an accurate, valid and appropriate assessment of the curriculum.

EYFS advisors need to have:

- substantial and recent expertise in early years settings
- a detailed understanding of the [EYFS framework](#)
- expertise in critically analysing assessment materials
- an understanding of the issues raised by recent reviews and consultations relating to assessment and how subsequent implementation may affect the future structure and strategy of assessments
- an understanding of key stage 1 (KS1), the adjacent key stage to EYFS
- the ability to communicate effectively with stakeholders, providing constructive feedback at meetings

6.2.1 Procurement process

EYFS advisors are appointed through an open competitive procurement exercise advertised on [Contracts Finder](#).

Applicants are required to respond to a series of questions to assess their skills and expertise. Responses are scored by relevant members of the assessment development research unit. Contracts are awarded to those with the highest overall

scores.

6.2.2 Contributions to assessment development

EYFS advisors typically review materials at the following stages of assessment development:

- initial drafts of items and mark schemes
- expert review 1, including participation in project board 1
- expert review 2, including participation in project board 2
- final assessment review meeting
- project board 3

The assessment development stages are summarised in [section 3](#).

6.3 Main areas for item review

The factors taken into account by STA and its reviewers are explained below:

6.3.1 Content domain

Items are reviewed to ensure the content domain, as detailed in the assessment specification, is applied accurately. This helps STA monitor the overall balance of content coverage in the assessments.

6.3.2 Assessment specification

Items and assessments are considered against the criteria laid out in the assessment specification, such as:

- item type
- cognitive domain
- context of questions

6.3.3 Construct irrelevant variance

Construct irrelevant variance (CIV) occurs when pupils interpret an item in an unintended way. As a result, their performance on the item measures something other than that intended by the assessment. This means that the item can become unexpectedly easy or difficult and does not contribute to the aim of the assessment.

Reviewing an item for CIV involves looking for unintended ways in which a pupil

might respond to or interact with the question and deciding whether these are valid misinterpretations of the item.

6.3.4 Distractors in selected response items

Selected response items, such as multiple-choice items, require careful review of distractors (incorrect answer options).

6.3.5 Accessibility of questions

All reviewers consider the needs of pupils with SEND and EAL, and some reviewers are specifically appointed for their expertise in one of these areas. Additionally, all items are reviewed by a modified test agency procured from the DfE modified framework that provides feedback on the suitability of the material for visually impaired and hearing-impaired pupils. STA applies the principles of universal design, where the standard version of the assessment is designed to be accessible to as many pupils as possible.

Accessibility settings and alternative versions of the assessment are developed for pupils with certain types of SEND (see [section 11](#)). However, a significant number of pupils with SEND will use the standard version of the assessments.

6.3.6 Other considerations

When considering the suitability of an item, reviewers also take account of:

- accuracy
- clarity of question text
- clarity of instructions
- appropriateness of contexts
- clarity and accessibility of illustrations
- layout of questions
- number of marks
- appropriateness of difficulty level
- compliance with house style
- digital interactions

7 Trialling

STA conducts trials with representative samples of pupils to assess whether items are valid assessments of pupils' knowledge and skills.

Co-ordination of the RBA trials is outsourced to approved suppliers. Trials take place in reception classes in the autumn term. Practitioners administer the RBA trials in schools using 2 digital devices to reflect the live administration. However, STA maintains overall management of the trials to ensure they are conducted in a robust, reliable and confidential manner.

Information from trialling is used to make decisions about whether an item enters the next stage of the assessment development process.

This section provides an overview of the trials.

7.1 Item validation trial

An IVT is the first large-scale trial in the assessment development process. The purpose of the IVT is to determine each item's suitability and general difficulty for the target age group and ability range. Qualitative and quantitative data is collected to inform the development of each item.

For the IVT, items are grouped into trial forms that reflect coverage in the live assessment. Each IVT item is administered to approximately 300 pupils. The sample of schools that participate in an IVT is as representative as possible of the characteristics of the national cohort in terms of school attainment and region.

There are 3 possible outcomes for each item following the IVT:

1. The item is progressed to the next stage of the development process with no amendments.
2. The item is progressed to the next stage of the development process with some amendments.
3. The item is archived because the evidence collected indicates that the item is not valid, reliable or of an appropriate level of difficulty and cannot be improved.

7.1.1 Item validation trial design

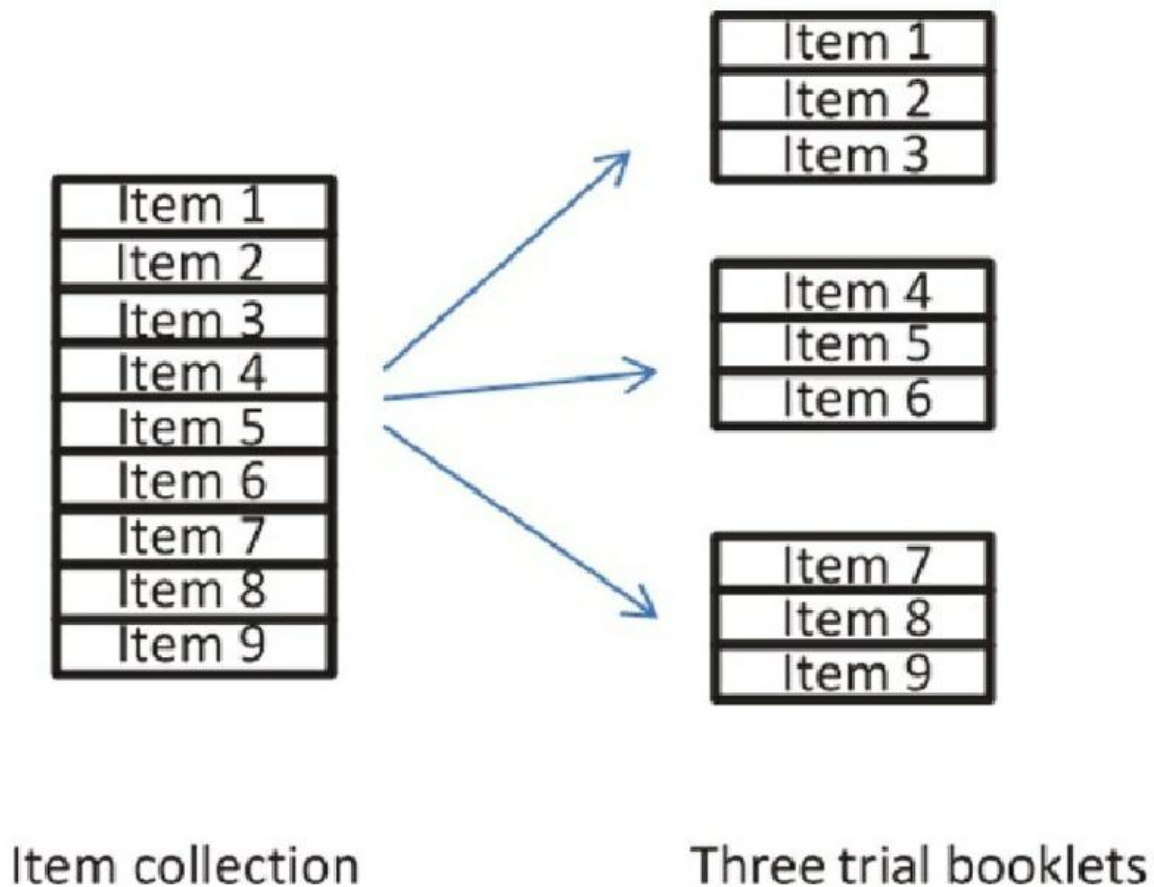
Not all items require an IVT, as certain item types and formats are more predictable in terms of performance and therefore go straight to a TPT.

IVT forms do not mirror the live assessment specification but are designed to

enable a large volume of items to be trialled.

An example of a trial design is presented in Figure 2, where a collection of 9 items is used to make 3 trial forms, with each item appearing once in the entirety of the trial:

Figure 2: IVT design



7.2 Technical pre-test

The TPT is the largest trial in the assessment development process. The purpose of the TPT is to gather detailed statistics to help support the final assessment construction process. Qualitative and quantitative data is collected about each item to determine suitability for a live assessment.

For a TPT, items are grouped into trial forms that reflect coverage in the live assessment specification. Each TPT item is administered to approximately 1,000 pupils. The sample of schools that participate in the TPT is as representative as possible of the national cohort of pupils, in terms of school attainment and region.

There are 3 possible outcomes for the items following a TPT:

1. The item is approved for use in a live assessment.
2. The item is not approved for use in a live assessment and is amended and re-trialled.
3. The item is not approved for use in a live assessment and is archived because the evidence collected indicates that the item is not valid, reliable or of an appropriate level of difficulty, and cannot be improved.

All items included in live assessments will have been subject to at least one TPT.

7.2.1 Technical pre-test trial design

The TPT trial design is more complex than that used for IVT. The same item appears in more than one form within the suite of trial materials. This helps to quantify the effect of any differences in overall ability between the groups of pupils completing the different forms.

Designing the trial so that items appear in more than one trial form also minimises the possibility that the performance of the item has been influenced by other items in the same form (item 'context'). The aim is that each item is trialled twice.

Figure 3: A sample technical pre-test design

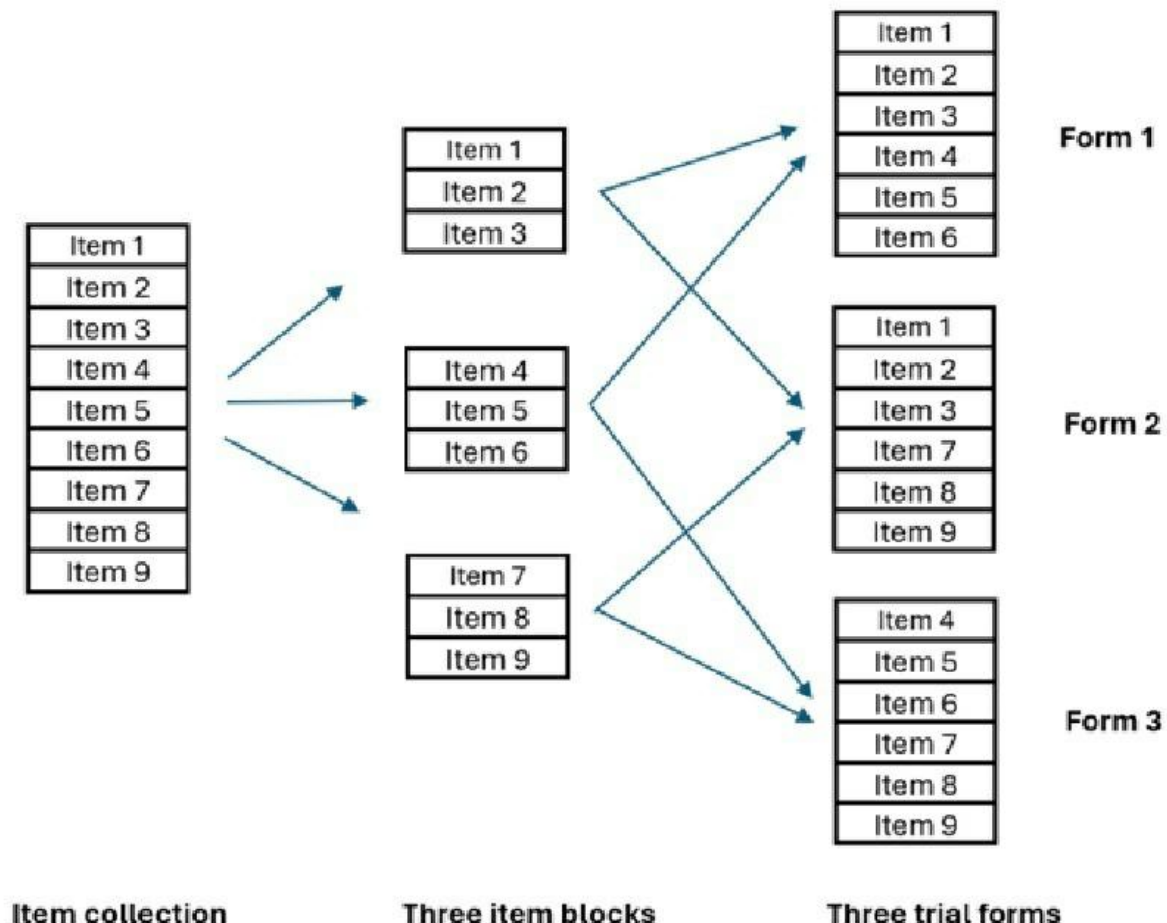


Figure 3 shows a simplified example of a TPT design, ensuring that each item

appears twice in the trial. A collection of 9 items is divided into 3 item blocks. The 3 blocks are used to make 3 trial forms by combining any 2 blocks. Items appear twice in the trial forms.

The trial design dictates that any form within the trial is attempted by 500 pupils, so that each item will be trialled by 1,000 pupils in total.

A TPT can vary in size depending on the number of items that need to be trialled.

7.3 Trial form construction

Trial forms for the IVT and TPT are constructed according to the chosen trial design and assessment specification for each subject.

For TPTs, statistical data gathered in the IVT is also reviewed to ensure the trial forms are of an appropriate difficulty level.

Within each assessment form, items are ordered by each substrand of the content domain and, where possible, in order of difficulty.

Forms are designed and scrutinised to ensure:

- there is no overlap of items that may prompt or support the answer to other items
- each contains an appropriate range of different item types, as set out in the assessment specification
- contexts that involve names of children include a mix of children of different gender and ethnicities

7.4 Conducting the trials

7.4.1 Trialling agencies

Trialling agencies procured from the DfE trialling agency framework facilitate the trials.

To be included on the framework, agencies must:

- meet security standards

- demonstrate sufficient resources and experience

7.4.2 Recruiting trial schools

From 2021, participation in trials for the RBA has been a statutory requirement for schools. The samples for the trials are designed to be representative of the national cohort.

Once the samples of schools have been recruited by the trialling agency, chi-square assessments for each stratifier are conducted to confirm the achieved samples meet STA's specification.

7.4.3 Trial administration

The trialling agency co-ordinates the trial administration by onboarding schools and providing them with helpdesk support. As administration is intended to mirror live assessment administration, school practitioners conduct the trials with the required number of pupils.

Questionnaires are used to collect qualitative information from practitioners about aspects of the assessment forms and their performance.

The trialling agency produces a final report for the trial. This is composed of:

- an administration report detailing the achieved trial samples, the representativeness of the samples as per STA's specification, and a summary of the trial
- a data report collating findings of the practitioner questionnaire data

7.5 Trialling data collection

During the administration of the trial, responses are captured in the system either by:

- the practitioner recording that the response was correct, incorrect or no response was given
- the pupil's interaction with the screen being automatically marked – for example, by selecting a multiple-choice option

The automatically marked responses are checked to confirm they are performing as expected according to the scoring rules.

During trialling, practitioners are required to enter verbatim responses for some

items which are also checked for accuracy and used by TDRs to iterate marking criteria. Timing and interaction data is also collected during trialling.

8 Analysis of trialling data

Statistical analysis of the items is carried out following each trial to inform the assessment development process. This analysis evaluates the performance of the items rather than the pupils.

From this analysis, STA gathers crucial, objective information about the difficulty of each item and how well it measures the assessment construct. It also highlights any potential flaws such as bias towards different groups of pupils or issues with the mark scheme. The types of analysis fall broadly under 2 areas of assessment theory:

- classical test theory (CTT)
- item response theory (IRT)

8.1 Classical analysis

Most CTT-based analysis is aimed at identifying how much information the observed score tells us about an individual's true score. A true score is a CTT concept interpreted as the average of observed test scores over an infinite number of administrations of the same assessment (per pupil). Classical analysis encompasses a range of statistics at item, form and assessment level.

Form-level or assessment-level information includes:

- score distributions
- measures of internal consistency reliability
- comparisons of mean scores by sex

Item-level statistics include:

- the mean score expressed as a percentage of marks
- the percentage of pupils achieving each possible score

- the percentage of pupils who omitted the item
- the percentage of pupils who omitted the item and all remaining items in the form

STA also examines the correlation between the item score and the total score, which has been corrected to exclude the item. This gives an indication of how well the item differentiates between pupils of differing ability.

8.2 Item response analysis

IRT refers to a group of statistical models in which the probability of pupil success on an item is a function of various item parameters, depending on the model. IRT allows the comparison of items that have not appeared in the same assessment together and pupils who have taken different assessments, by putting them all onto a single common scale, linked by common items or pupils. Due to its complexity, IRT is correspondingly more difficult to carry out, requiring specialist software.

IRT assumes there is an underlying latent construct being measured that is continuous in nature. An item characteristic curve plots the probability of a correct response as a function of the latent construct. Item characteristic curves are combined to plot an assessment characteristic curve for each assessment.

Another important concept in IRT is that of item information. This provides a display of item contribution to the latent construct and depends to a large extent on the differentiation power of the item along the latent construct continuum.

IRT has a number of assumptions that must be statistically tested to ensure its use is appropriate – that the:

- items in the assessment fit the IRT model
- underlying latent construct is unidimensional
- items in the assessment are independent of each other

8.3 Distractor analysis

Numeric codes are assigned to represent how the pupil responded to the options in multiple-choice items. Distractor analysis is used to analyse the response codes to

provide more information about how the pupils answer the item – for example, showing how often incorrect options were selected.

8.4 Differential item functioning

Differential item functioning (DIF) is a statistical analysis carried out to flag differences in item performance based on group membership. This analysis is useful when looking at all of the evidence for assessment construction. However, DIF can only indicate that there is differential item performance between groups (boy and girl, EAL and non-EAL) that have the same overall performance. It cannot determine the cause of the differential performance.

Further qualitative exploration of an item leading to a reasoned, substantive explanation for the DIF is required before an item could be considered biased. It is important to acknowledge that group differential performance does not, on its own, indicate bias or lack of fairness.

8.5 Timing and interaction data

As the RBA is administered digitally, a range of timing and interaction data is captured during trialling. Psychometricians provide analysis of this data for TDRs to evaluate alongside the other data sources to inform item refinement. This analysis can be tailored to meet specific questions TDRs might have for particular items or groups of items – for example, how many drags or drags between drop zones were made by pupils in sequencing item types.

8.6 Analysing qualitative data

Qualitative evidence is gathered throughout the trial administration. Feedback is evaluated alongside the quantitative data and used to inform item refinement.

8.6.1 Summary of types of evidence gathered

Practitioners whose pupils participate in the trial complete a questionnaire. The questionnaires provide an opportunity for practitioners to give their opinions on the suitability of the items in terms of accessibility, clarity and layout, manageability and

timing.

8.6.2 Revising items and coding frames

Qualitative data can support information provided by quantitative data and highlight issues with accessibility or reasons why pupils may have misinterpreted a question.

However, because it is subjective and because of the relatively small numbers responding to the questionnaires, some caution is required when using qualitative data to make judgements on whether changes should be made to items.

9 Assessment construction

Assessment construction is the process by which items are put together into assessment forms that meet the specification set out in the assessment framework (see [section 2](#)).

TDRs and psychometricians are responsible for constructing the assessments based on the evidence from trialling. After the live assessment has been constructed, it is reviewed at expert review 3 and by EYFS advisors to ensure the assessment is fit for purpose and meets the assessment specification.

9.1 Item finalisation

All questions considered for a live assessment have been through a TPT, which has provided evidence that they are functioning well. Items going forward to the live assessment should not be amended after the TPT, as this could affect pupil performance on the item and render the data held unreliable.

An item finalisation meeting is held following each trial. At this meeting, TDRs and psychometricians confirm which items from the trial are suitable for inclusion in the next round of trialling or in a live assessment. Items that are deemed suitable must:

- have appropriate facilities (difficulty levels)
- differentiate between groups of pupils of differing ability
- be accessible in terms of language, layout and illustrations

Both quantitative and qualitative data from the trial is reviewed to give a report on

each item's reliability and validity.

Any items that do not function well, or that had poor feedback from teachers or pupils, are either removed from the pool of available items for selection or amended to be re-trialled for future use.

Occasionally, it is not possible for the items in the live assessment to be identical to the trialled TPT versions. Small changes may be required at the assessment construction phase – for example, to keep the gender and ethnicity of pupils named in the papers balanced.

While there is a small chance this could affect the performance of a question in the live assessment compared to its performance in the TPT, it is unlikely to be significant and is monitored through the live analysis. Such changes are only made when essential.

9.2 Constructing the assessment

STA's psychometricians use question and item metadata to construct a draft assessment that meets the assessment specification and optimises the measurement precision within the appropriate range on the ability scale. An assessment construction meeting is then held to select the items for IVT, TPT or live assessment forms. At this meeting, TDRs use the psychometrician's initial selection as a starting point. The meeting's participants consider the proposed forms, taking into account:

- item type
- presentational aspects
- question contexts and coverage
- whether there are any conflicts between what is assessed within assessment forms and across the assessment as a whole

At this stage, items may be swapped in or out of the assessment to improve its overall quality and suitability.

Once the items have been chosen for each assessment, they are put in an appropriate order within each form. The items are placed in order of difficulty, where possible, within each substrand of the content domain.

9.3 Reviewing and confirming the assessment

The assessment is subject to a further review process after assessment construction. It is reviewed first by TDRs and the psychometrician to ensure none of the assessment constraints have been missed during the assessment construction meeting.

External stakeholders then review proposed live assessments at expert review 3.

Live assessments are then presented to project board 3 (see [section 10](#)). Once the project board is satisfied that the assessment has met the assessment specification and is of sufficient quality, the assessment is signed off to publish.

9.4 Quality assurance and proofreading

The assessment goes through a rigorous quality-assurance process before the materials are signed off to publish.

9.4.1 Proofreading

At each stage of proofing, a range of people proofread the assessment to ensure all the material is accurate and free from errors. Between each stage, any errors found are corrected and rechecked.

The assessment is proofread by people with particular areas of expertise. They make sure that:

- there are no typographical errors
- house style is applied consistently
- grammar is correct
- there are no content overlaps
- the assessment gives pupils all the information they need to answer the questions
- assessment content is factually correct and would stand up to scrutiny
- the number of marks is correct
- ISBN numbers, product codes and barcodes are correct where they are used (for paper resources)

- there is consistency across different versions of the forms

TDRs also proofread the materials at every stage and have overall responsibility for the quality of the assessments. They are responsible for collating the comments from each proofing round and ensuring any amendments are made.

9.4.2 Business acceptance testing

Business acceptance testing (BAT) is carried out internally to verify that the assessment's digital functionality operates as intended within the assessment service. The TDRs lead and co-ordinate a series of testing activities that assess key elements of the assessment functionality. These include:

- completing a series of partially scripted assessments to enable psychometricians to validate data outputs
- reviewing all assessment items with accessibility settings
- completing assessments using the paper-based materials
- completing assessments with scripted responses to test routing logic
- checking the accuracy of narrative statements generated by a series of scripted assessment responses
- having assessments completed by internal users with no prior experience of the assessment service to identify usability or comprehension issues

Any issues identified during these activities are shared with the relevant team for remediation. Once the fixes are implemented, testing is repeated to quality-assure the remediation work.

9.4.3 Digital services testing

The assessment service has attained Authority to Operate (AtO), enabling it to be used in a live government environment. This confirms that the service meets required security, risk, and operational standards, supported by a range of internal and external assurance activities.

These include:

- annual IT health check and penetration testing – independent security assessments that audit the service and identifies potential vulnerabilities or issues, ensuring risks are identified promptly and appropriate remediation actions are taken to maintain the security of the service
- Government Digital Service (GDS) assessments – formal reviews measuring the service against the Service Standards, including user needs, accessibility,

security, agile delivery, and reliable operation

- the Government Internal Audit Agency (GIAA) audit – an independent audit of the systems, testing processes and controls in place within the service, providing assurance over governance, risk management, and operational effectiveness

10 Governance

Overall governance of the RBA is managed through fortnightly checkpoint meetings involving the whole project team, plus relevant members of STA's senior management team. Outcomes of these meetings are reported to monthly delivery and technical subprogramme boards as appropriate, chaired by a deputy director and attended by members of STA's senior management team. Escalation is through normal STA routes, including the Executive Management Board.

In addition, project boards take place at 3 points in the assessment development process:

- before IVT
- before TPT
- before the materials go live

The purpose of the project board is to:

- review documentation on the quality of the items and assessments
- ensure there is sufficient material available to carry on to the next stage of development
- ensure, for assessments going live, that the correct process has been followed, the assessment meets the specification and the materials are fit for purpose

The project board will consider the evidence carefully before approving the assessment materials for trial or live assessment.

Statutory national testing is regulated by Ofqual, whose representatives may observe technical or project boards.

10.1 Project board 1

Project board 1 takes place shortly after the resolution meeting for expert review 1 (for a summary of the assessment development process, see [section 3](#)). The purpose of the meeting is to formally approve materials proceeding to the IVT.

At project board 1, the following evidence is considered:

- trial forms
- a summary of the process to this point
- an overview of evidence from expert review 1 and any feedback from the EYFS advisors on the materials
- coverage of items in the trial against the assessment framework
- coverage of items within the item bank to ensure areas not covered in the trial are provided for in the item bank

10.2 Project board 2

Project board 2 takes place shortly after the resolution meeting for expert review 2. The purpose of the meeting is to formally approve materials for inclusion in the TPT.

At project board 2, the documentation comprises:

- trial forms
- an outline of the TPT model being used
- a summary of the process to this point
- an overview of evidence from expert review 2 and the EYFS advisors
- coverage of materials in the trial against the assessment framework
- coverage of items within the item bank to ensure areas not covered in the trial are provided for in the item bank – in total, the materials in the trial and the item bank must be able to produce a live assessment that meets the specification

10.3 Project board 3

Project board 3 takes place shortly after expert review 3. The purpose of project board 3 is to formally approve the materials going forward as the live assessment.

Any recommendations for changes to the constructed assessment following expert review 3 must be made at project board 3.

Project board 3 reviews the validity evidence for the assessment, including whether the correct process has been followed. Project board 3 ensures that the proposed assessment meets all aspects of the specification. If these conditions are met, the materials can be approved as fit for purpose.

11 Assessment administration

STA is responsible for developing assessment guidance and managing the processes related to assessment administration and access arrangements.

11.1 Assessment access

Maintained schools and academies, including free schools, special schools and Defence Children Services (DCS) schools and settings with a reception cohort will be able to access the 2 assessment services required to administer the RBA (see [section 3.2](#)).

Schools can request paper assessment materials for pupils who require them by contacting STA using the [RBA help centre](#).

11.2 Pupil registration

All pupils enrolled in reception at maintained schools, special schools and academies, including free schools, must be registered for the assessment. This includes any pupils who are unable to access the assessment.

11.3 Assessment materials

Assessment materials are available to schools ahead of the live delivery period and practitioners are able to preview them in advance of administration. Headteachers ensure the security of the assessments is maintained. Schools follow the instructions in the [RBA assessment and reporting arrangements](#) on keeping assessment materials secure and treat them as confidential from the point they are received in school until the end of the relevant assessment administration period.

11.4 Administering the assessments

The RBA must be administered to participating pupils in the first 6 weeks after they enter reception. Assessments must continue throughout the academic year for any pupils who join reception and have not already participated in the RBA.

11.5 Communications and guidance

STA publishes [RBA guidance](#) throughout the assessment cycle to support schools with the assessment process. The [RBA administration guidance](#) is developed to ensure consistency of administration across schools and therefore supports the validity evidence in relation to standardisation. Additional guidance is also available in the assessment services.

11.6 Access arrangements

Pupils with specific needs may need additional arrangements so that they can take part in the assessment. Access arrangements are adjustments that can be put in place to support pupils and ensure they are able to fully demonstrate their attainment. This supports STA in determining that the assessments are fair for all groups of pupils. Headteachers and teachers must consider whether any of their pupils will need adjustments before they administer the assessments.

Access arrangements must never advantage or disadvantage the pupils. Any

support should be based primarily on usual classroom practice. The support given must not change the assessment items and the answers must be the pupil's own. Failure to apply access arrangements appropriately can result in a maladministration investigation at the school, and pupils' results may be amended or annulled.

The [RBA modified administration guidance](#) advises on the range of arrangements available to help pupils with specific needs access the RBA. These can be used to support pupils:

- with a hearing impairment
- with a visual impairment
- who use sign language
- who have difficulty concentrating
- who have processing difficulties

11.6.1 RBA administration

The RBA has been designed to be accessible to the vast majority of pupils on entry to reception. The one-to-one administration approach allows practitioners to adjust their delivery of the assessment to support the needs of individual pupils. The practitioner can:

- move through the assessment at a pace that suits the pupil, as there is no time limit
- pause the assessment to take breaks as required
- support the pupil to interact with their touchscreen

11.6.2 Accessibility settings

Accessibility settings are available in the assessment service. They can be viewed and managed by the practitioner to adapt the appearance of the standard assessment if required. Settings can be used to:

- change the background colour
- add borders to numbers
- change the colour contrast of numerals
- change images to simplified or greyscale versions

11.7 Paper assessment

If a pupil cannot access a touchscreen, a paper version is available which allows the assessment to be administered using the practitioner device and paper assessment materials for the pupil, or using paper materials for both.

12 Assessment validity

Statutory national assessments are regulated by Ofqual. Ofqual's statutory objectives are to promote standards and public confidence in national assessments, and its primary duty is to keep all aspects of national assessments under review. Ofqual focuses on the validity of assessment and takes a risk-based approach: observing, scrutinising and reporting on key aspects of assessment arrangements relating to validity.

The [RBA validity framework](#), which is provided as an appendix to this handbook, sets out the evidence that STA has generated in relation to the expectations in Ofqual's regulatory framework for national assessments. The regulatory framework focuses on validity, including the concepts of reliability, comparability, minimising bias and manageability.

12.1 Validity argument

The purpose of the RBA is to create a baseline measure from which cohort-level progress to the end of KS2 can be calculated. To demonstrate that the assessment developed meets this purpose, STA has determined 4 claims that must be satisfied:

- Claim 1: The assessment is representative of a range of literacy, communication and language (LCL) and mathematics skills and knowledge appropriate to the age and development of children at the start of reception
- Claim 2: Assessment results provide a fair and accurate measure of pupil performance
- Claim 3: Pupil performance is comparable within and across schools

- Claim 4: The meaning of total scores is clear to those responsible for creating the progress measures [The meaning of assessment scores is clear to stakeholders]

STA developed a number of questions for each claim, to enable evidence to be provided that the claim has been met.

13 Glossary

Phrase	Definition
Chi-square test	a statistic, testing the relationship between the row and the column in a table
Cognitive domain	thinking and processing skills specific to an assessment subject
Content domain	the assessable elements of the programme of study
Differential item functioning (DIF)	an analysis which indicates whether different groups of pupils at the same level of attainment have the same probability of a correct response on an item
Distractor	the incorrect options given for a multiple-choice item
Inter-rater	consistency between 2 or more judges
Item	within an assessment, a 'question' is referred to as an item – the lowest level to which a mark can be awarded
Item bank	database holding item, assessment and process information
Item classification	metadata collected on item characteristics to ensure assessments have appropriate coverage over time
Mark scheme	a document outlining the requirements for awarding marks on an item level
Observed	the reported score for a pupil

score

Reliability	indicates the dependability, consistency, or freedom from random measurement error of an assessment score
Selected response item	an assessment question that requires pupils to select their answer from one of several possibilities given
Assessment framework	subject-specific documents containing detail to guide the assessment development process
True score	a classical test theory concept interpreted as the average of observed assessment scores over an infinite number of administrations of the same assessment (per pupil)
Validity	the degree to which evidence gathered in the assessment development process supports the assessment outcome interpretation that the assessment measures what it is intended to measure

1. The Standards for Educational and Psychological Testing (2014) were prepared by the joint committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. [↩](#)

[↑ Back to top](#)

Help us improve GOV.UK

To help us improve GOV.UK, we'd like to know more about your visit today. [Please fill in this survey \(opens in a new tab\)](#).





Services and information

[Benefits](#)

[Births, death, marriages and care](#)

[Business and self-employed](#)

[Childcare and parenting](#)

[Citizenship and living in the UK](#)

[Crime, justice and the law](#)

[Disabled people](#)

[Driving and transport](#)

[Education and learning](#)

[Employing people](#)

[Environment and countryside](#)

[Housing and local services](#)

[Money and tax](#)

[Passports, travel and living abroad](#)

[Visas and immigration](#)

[Working, jobs and pensions](#)

Government activity

[Departments](#)

[News](#)

[Guidance and regulation](#)

[Research and statistics](#)

[Policy papers and consultations](#)

[Transparency](#)

[How government works](#)

[Get involved](#)

[Help](#) [Privacy](#) [Cookies](#) [Accessibility statement](#) [Contact](#)

[Terms and conditions](#) [Rhestr o Wasanaethau Cymraeg](#)

All content is available under the [Open Government Licence v3.0](#), except where otherwise stated

