



Standards
& Testing
Agency

Reception baseline assessment validity

Report on the RBA pilot

February 2020

Contents

Table of figures	3
Summary	4
Who is this publication for?	4
The assessment	5
The online system (BeP)	6
Sources of evidence	7
RBA pilot	7
Sample representation for schools in the pilot	8
Pupil numbers and exclusions from the pilot analysis	11
Analysis methodology	12
Validity framework	13
Claim 1	14
Research questions for claim 1	14
Claim 2	22
Research questions for claim 2	22
Claim 3	30
Research questions for claim 3	30
Claim 4	33
Research questions for claim 4	33
References	34

Table of figures

Table 1: Summary of participation in the pilot.....	8
Table 2: Representation of the sample used in analysis by government office region.....	9
Table 3: Representation of the sample used in analysis by school type.....	9
Table 4: Representation of the sample used in analysis by reading attainment	10
Table 5: Representation of the sample used in analysis by writing attainment.....	10
Table 6: Representation of the sample used in analysis by mathematics attainment.....	11
Table 7: Number of pupils with complete and discontinued assessments overall.....	12
Table 8: Number of pupils with complete and discontinued mathematics component.....	12
Table 9: Number of pupils with complete and discontinued LCL component.....	12
Table 10: Supporting research evidence for mathematics content domain	17
Table 11: Supporting research evidence for LCL content domain	18
Table 12: Approximate proportion of assessment items by mathematics component	20
Table 13: Approximate proportion of assessment items by LCL component.....	21

Summary

The reception baseline assessment (RBA) is a new national assessment of literacy, communication and language (LCL) and mathematics that is intended to be administered in reception classes in all primary, infant and first schools in England from 2020. It is intended to form the baseline for the primary progress measure, ensuring that the progress made by pupils throughout their time in primary school is recognised.

The National Foundation for Educational Research (NFER) is contracted by the Department for Education (DfE) to develop, trial and pilot the new RBA from May 2018, and deliver it from September 2020.

This report provides non-statutory guidance from the Department for Education on the validity evidence gathered throughout every stage of the development of the RBA. It has been produced to help those with an interest in assessment to understand the validity argument that supports the RBA.

Who is this publication for?

This document 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.

The assessment

The reception baseline has been designed to be an age-appropriate assessment to be taken by children during their first half term in reception (the academic year in which they turn five). It assesses two components: mathematics and LCL. The assessment consists of practical tasks for the child and makes use of physical resources. Administration instructions for each task and the recording of the assessment are provided via the BeP. This aims to ensure that the assessment is delivered and scored consistently across all schools.

An administration guide, training videos for four tasks, and a preview assessment are provided to support practitioners in the delivery of the RBA. Those conducting the assessment should be fully trained and familiar with these materials before administering the assessment and headteachers are required to sign a declaration to confirm that this is the case.

The assessment must be administered by a reception teacher, reception teaching assistant or other suitably qualified practitioner (for example, the early years lead or SENCo), working one-to-one with each child. The practitioner should be familiar to the child.

Responses are recorded by the practitioner using the BeP and scoring is calculated automatically. Each item requires a single, objective, binary yes/no scoring decision to be made. The assessment includes carefully designed routing which helps to prevent pupils from being presented with too many activities in which they are unlikely to be successful. It also helps to reduce the time required for the assessment and the possible loss of motivation that pupils may feel if they are unable to complete an activity. The routing rules are automatically applied by the online recording system. Practitioners do not receive any numerical scores, and there is no pass mark. However, individual reports containing narrative statements related to the outcomes of the assessment for each child can be downloaded from the BeP.

The reception baseline has been designed to be an inclusive assessment, accessible to the majority of pupils on entry to school. It has been designed so that children with Special Educational Needs and Disabilities (SEND) and those learning English as an additional language (EAL) can participate in the standard assessment format and has been subject to SEND and cultural reviews.

Modified resources are available for children with a visual impairment and practitioners are also able to make further adaptations to make resources appropriate for their pupils. For children with a hearing impairment or who use sign language, the assessment can be conducted in British Sign Language or any sign-supported English, using signs familiar to the child.

The BeP has a number of functions that support delivery of the RBA. At any point during the administration, the practitioner can pause the assessment if it needs to be temporarily stopped or if the child needs a break. These pauses can be any length in duration and the practitioner is asked to record a reason for pausing the assessment. The practitioner also has the option to discontinue the assessment component if they feel that it is not appropriate to continue, for example, if the child is having difficulty in accessing the content. In a minority of cases, it may be necessary for a child to be disapplied if a practitioner believes that the child (for example, a child with complex SEND) will not be able to access any element of the assessment.

The online system (BeP)

The RBA is carried out by practitioners using the BeP. This online system is split into multiple areas, which include:

- a training area where the training videos and the preview assessment are accessed
- an area to upload and manage pupil data
- an area to add and manage staff accounts
- an assessment area, from where pupil assessments are launched and the assessment status for each pupil can be reviewed
- an area to access reports for pupils who have completed the RBA
- the headteacher's area, where the headteacher declaration form can be accessed

In addition to the materials provided to support the administration of the assessment, schools also have access to a helpline which is available via email or by telephone.

Prior to the start of the pilot, the BeP was thoroughly reviewed, including all aspects relating to security of the system, and was approved by the DfE for use in schools for the pilot year.

Sources of evidence

The evidence used in this report is drawn from a number of sources:

- The development and implementation of the NFER optional reception baseline (2015)
- Reviews by experts during the development of the RBA
- The trial of the RBA in the academic year 2018-19
- The pilot of the RBA in the academic year 2019-20

RBA pilot

The pilot enabled evaluation of the effectiveness of individual questions, as well as the assessment as a whole; of the functionality of the Baseline ePortal (BeP) online system; of the logistical processes involved; and of school experience. It also allowed for the identification of any recommendations for change.

Whilst participation is voluntary, the pilot is a key part of the development process, which will be used to ensure that:

- the assessment approach, systems and guidance are fit for purpose
- the outcomes of the assessment meet all key requirements, including ensuring that the assessment is reliable and appropriate for use in the first six weeks of reception

All maintained infant, first and primary schools across England, including special schools and service children's education schools, were invited to participate in the pilot phase of the RBA. Independent schools and pupil referral units were excluded. Sign up took place within a five-week window, 1 March – 3 April 2019.

The pilot year of the RBA commenced in August 2019 and will continue until the end of the summer term 2020. This report is based on data provided by the 8,994 schools who uploaded pupil data to the BeP during the first half term of the pilot year (up to and including 25 October 2019), during which the majority of children in this cohort had been assessed.

A summary of the number of schools signing up and participating in the pilot, up to and including 25 October 2019, is provided in the table below.

	Number of schools
Initially signed up to RBA pilot	9,657
Total number of withdrawals	415
Total number of schools logging in to the assessment system (up to and including 25 October 2019)	9,128
Total number of schools uploading pupil data to the assessment system (up to and including 25 October 2019)	8,994
Total number of schools completing assessments for all uploaded pupils (up to and including 25 October 2019)	7,046

Table 1: Summary of participation in the pilot

Note: There are instances where a school may have initially engaged and then subsequently withdrawn from the pilot. This is the reason why there is some overlap between the numbers.

Sample representation for schools in the pilot

Of the 9,128 schools who logged into the system in the pilot, 7,046 completed assessments for all pupils during the first half of the autumn term.

Analysis was conducted to determine if the schools that uploaded pupil data to the BeP) were representative of the national school population in terms of location, school type and key stage 1 attainment in reading, writing and mathematics (measured as the proportion of pupils reaching the expected standard or higher). The significance tests performed on this whole data set indicated that the data was not representative as it had statistically significant differences from the national school population both geographically and in terms of school type.

Since the clearest picture of the assessment's functioning can be gained through the use of a representative dataset, a sub-sample of all pilot schools was drawn such that it would be representative of the national school population on all variables of interest.

This resulted in a sample of 4,046 schools being drawn for the analysis. The composition of this sample and that of the national population is presented in tables 2 to 6 below.

Government office region	Number of schools in sample	Percentage of schools in sample	Number of all schools	Percentage of all schools
East Midlands	355	8.8	1,518	9.4
East of England	488	12.1	1,857	11.5
London	441	10.9	1,733	10.8
North East	224	5.5	839	5.2
North West	600	14.8	2,393	14.9
South East	602	14.9	2,337	14.5
South West	462	11.4	1,802	11.2
West Midlands	440	10.9	1,702	10.6
Yorkshire and the Humber	434	10.7	1,716	10.7
Missing ¹	0	0	184	1.1
Total	4,046	100	16,081	100

Table 2: Representation of the sample used in analysis by government office region

Note: percentages may not sum to 100 due to rounding.

School type	Number of schools in sample	Percentage of schools in sample	Number of all schools	Percentage of all schools
Infants	320	7.9	1,214	7.5
Infants and Juniors	3,693	91.3	14,544	90.4
All through	33	0.8	136	0.8
Missing	0	0	187	1.2
Total	4,046	100	16,081	100

Table 3: Representation of the sample used in analysis by school type

Note: percentages may not sum to 100 due to rounding.

¹ The data used in this and subsequent tables comes from a variety of sources that were generated at different points in time. These sources needed to be matched in order to undertake analysis and this was not possible in a small number cases (for example, where schools had changed their unique reference number (URN) between the two data collection exercises). Where data could not be matched, the schools are recorded as having 'missing' data in these tables.

Reading attainment band (key stage 1 teacher assessment) ²	Number of schools in sample	Percentage of schools in sample	Number of all schools	Percentage of all schools
Lowest quintile	777	19.2	3,019	18.8
Second lowest quintile	761	18.8	3,087	19.2
Middle quintile	778	19.2	3,068	19.1
Second highest quintile	795	19.6	3,138	19.5
Highest quintile	815	20.1	3,060	19.0
Missing	120	3.0	709	4.4
Total	4,046	100	16,081	100

Table 4: Representation of the sample used in analysis by reading attainment

Source: ONS

Note: percentages may not sum to 100 due to rounding.

Writing attainment band (key stage 1 teacher assessment)	Number of schools in sample	Percentage of schools in sample	Number of all schools	Percentage of all schools
Lowest quintile	758	18.7	3,047	18.9
Second lowest quintile	737	18.2	3,025	18.8
Middle quintile	914	22.6	3,430	21.3
Second highest quintile	676	16.7	2,781	17.3
Highest quintile	841	20.8	3,089	19.2
Missing	120	3.0	709	4.4
Total	4,046	100	16,081	100

Table 5: Representation of the sample used in analysis by writing attainment

Source: ONS

Note: percentages may not sum to 100 due to rounding.

² This work contains statistical data from ONS which is Crown Copyright. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Mathematics attainment band (key stage 1 teacher assessment)	Number of schools in sample	Percentage of schools in sample	Number of all schools	Percentage of all schools
Lowest quintile	751	18.6	3,048	19.0
Second lowest quintile	797	19.7	3,055	19.0
Middle quintile	809	20.0	3,146	19.6
Second highest quintile	772	19.1	3,191	19.8
Highest quintile	797	19.7	2,932	18.2
Missing	120	3.0	709	4.4
Total	4,046	100	16,081	100

Table 6: Representation of the sample used in analysis by mathematics attainment

Source: ONS

Note: percentages may not sum to 100 due to rounding.

Pupil numbers and exclusions from the pilot analysis

After drawing the sample of schools for the analysis, checks were made on the viability of the pupils in this sub-sample. Upon inspection of this data, a number of exclusions were made:

- pupils with inappropriate/out of range dates of birth (those recorded as being before 1 April 2014 or after 31 August 2015)
- pupils missing data due to system outages
- pupils who were not presented with all the necessary questions³

At the component level, only those pupils deemed to have finished the component (completed/discontinued) were included in the analysis. For the overall assessment analysis, only those who had finished both components were included. As a result, the total number of pupils for each component is different since some pupils had completed one component but not the other.

These considerations yielded the samples described in tables 7 to 9 below.

³ There were a small number of incidents during the pilot where system issues resulted in missing data for 7,285 pupils, which equates to 2% of pupils who were involved in the pilot. These issues have been resolved and fixes have been implemented to ensure they do not reoccur during live administration.

Overall status	Pupils
Mathematics complete / LCL complete	152,080
Mathematics complete / LCL discontinued	84
Mathematics discontinued / LCL complete	158
Mathematics discontinued / LCL discontinued	180
Total	152,502

Table 7: Number of pupils with complete and discontinued assessments overall

Mathematics status	Pupils
Complete	154,526
Discontinued	306
Total	154,832

Table 8: Number of pupils with complete and discontinued mathematics component

LCL status	Pupils
Complete	154,809
Discontinued	397
Total	155,206

Table 9: Number of pupils with complete and discontinued LCL component

Analysis methodology

Data from the representative sub-samples were analysed to establish the outcomes of the two assessment components: mathematics (n=154,832) and LCL (n=155,206). For pupils in these sub-samples that had completed both the mathematics and LCL components (n=152,502), performance on the whole assessment was also analysed.

Assessment data was analysed using both Classical Test Theory and Item Response Theory to derive a number of measures including item level functioning (for example, facility and discrimination values), differential item functioning and overall reliability. Consideration was given to each of the two components, mathematics and LCL, and to the assessment as a whole.

Analysis of the time taken to complete the assessment was also performed.

Validity framework

The purpose of the RBA is to identify pupils' starting points in order to create reception to end of key stage 2 school level progress measures. To demonstrate that the assessment meets this purpose, STA has determined four claims that must be satisfied:

- Claim 1 – the assessment is representative of a range of 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

STA developed a number of questions for each claim to enable evidence to be provided that the claim has been met. The remainder of this report sets out these claims and related questions and provides the evidence to support the argument that the claim has been met.

Claim 1

The assessment is representative of a range of LCL and mathematics skills and knowledge appropriate to the age and development of children at the start of reception.

Research questions for claim 1

1.1 Are the assessable areas of LCL and mathematics clearly defined as a content domain?

The content domain for the RBA is defined in the [Assessment framework – Reception Baseline Assessment](#) (page 6). A draft of the assessment framework was published in February 2019 and a final assessment framework was published in February 2020.

The content domain was developed by NFER’s assessment researchers in consultation with STA’s test development researchers. The content domain was based on the early years outcomes in the early years foundation stage (EYFS) and also looks to later outcomes at key stages 1 and 2.

The RBA reflects aspects of the learning and development requirements of the EYFS. However, due to the length and nature of the reception baseline, not all areas of the EYFS are assessed. The assessment samples from a range of LCL and mathematical knowledge and skills appropriate to the age group.

The assessment consists of:

- mathematics tasks
 - early number
 - early calculation (early addition/subtraction)
 - mathematical language
 - early understanding of pattern
- LCL tasks
 - early vocabulary
 - phonological awareness
 - early comprehension

1.2 Are the areas of the content domain an accurate reflection of the range of skills and knowledge of children of this age and stage of development?

The RBA is intended to be short, taking approximately 20 minutes per pupil, and assess aspects of mathematics and LCL suitable for children of this age. The RBA content domain samples from the range of elements of early learning, as set out in the EYFS documentation *Development Matters*. Tasks assessing skills associated with different ages and stages of development are included in the assessment. These range from the statements that indicate typical development for children aged from 22 months to the

early learning goals that are used to assess pupils at the end of the Reception year. Information on how content was sampled is provided below.

It is necessary for the assessment to be suitable for the full range of attainment that can be seen with this age group. This is to ensure that it reflects the variety of experience that different children will bring with them to reception. In order to avoid a ceiling effect, where children cannot demonstrate their actual attainment because the tasks being presented are too easy, it is important to include more challenging tasks. However, as it is not the expectation that all children will be able to successfully answer all questions, the assessment is routed. This means that children should not be presented with a significant number of activities that are too difficult for them.

The purpose of the RBA is to form a starting point from which the DfE can measure progress. Early development work involved establishing appropriate content domains for the RBA and included reviewing existing research into which aspects of mathematics and LCL have been shown to be related to later attainment to ensure that the most suitable content of the EYFS was sampled in the assessment. A summary of the research is provided on pages 16 to 18.

The practitioner panel reviewed all of the content domains and confirmed that they were appropriate. They further confirmed that the associated assessment tasks were appropriately allocated.

Feedback from the 2018 trial of the RBA (practitioner feedback $n=313$) indicated that the majority of practitioners considered that the mathematics and LCL tasks were suitable for use with children at the start of reception, with 82% of practitioners considering the mathematics tasks to be at least 'satisfactory' and 65% of practitioners rating the LCL tasks as 'satisfactory' or above. In order to ensure that only the most appropriate items are used in the RBA, more items were trialled than would be needed in the final assessment. The final RBA contains about half of the items that were trialled. Specific comments about tasks received in the feedback from the trial were considered during the item selection process, helping to ensure that only those items considered most appropriate were included.

Feedback from the first half term of the pilot (practitioner feedback $n=2,507$) indicated that the majority of practitioners felt that the assessment tasks were appropriate for children in reception. In total, 84% of practitioners rated the children's interest and enjoyment of the tasks as at least 'satisfactory'. An even higher proportion, 89%, said that the children's understanding of the tasks was 'satisfactory' or better.

The tasks were also reviewed by a practitioner panel and an expert panel (both convened by NFER). Both panels provided feedback indicating that the tasks were appropriate for children of this age group.

Following the pilot, STA made the decision to remove one LCL task and one mathematics task from the assessment. The LCL task related to early reading, which

very few pupils completed due to routing and even fewer pupils answered correctly. The mathematics task related to shape. Although this task functioned appropriately and was answered correctly by many pupils, it was removed to balance the assessment between the two components.

These changes will have the effect of reducing the time required to complete the assessment without compromising the quality of the assessment.

Following these changes, the data from the pilot supports the assertion that the assessment was age-appropriate. There is a good spread of pupils across the score range, with less than 0.8% of pupils scoring no marks and less than 0.4% achieving full marks. This provides evidence that there is not a ceiling effect on the assessment and that it can discriminate well between pupils across the ability range.

Not reached rates were calculated as the percentage of pupils who were not presented with an item because the assessment was discontinued. During the pilot, this was very rare, with only 0.2% of pupils having been stopped before the end of the maths component and only 0.3% of pupils having been stopped before the end of the LCL component.

Analysis shows that the routing worked effectively to ensure the demand of the assessment met the needs of different pupils.

The online system recorded how long it took to complete each task as well as the length of time taken to complete the assessment overall. Once the changes to the assessment outlined above are made, analysis of timing data shows that, on average, the assessment will take just over 14 minutes to complete, with 92% of administrations being completed within 20 minutes.

Content domain: Mathematics

Competence in early mathematics is strongly related to later school success. The relationship between early number competence and later mathematical achievement has been well established (Aubrey *et al.* 2006; Jordan *et al.*, 2009). There is also evidence that competency in counting, and in particular the development of higher-level counting skills, predicts later mathematical achievement, and having a solid understanding of counting principles will increase pupils' ability to undertake more complex mathematics tasks later in their school careers (Nguyen *et al.*, 2016). The development of the RBA mathematics tasks was based on research evidence, including research that examined good predictors of success or difficulty in this area.

A summary of the mathematics content domains, and supporting research that underpins their development, is provided in the table below.

Content domain	Supporting research evidence used to inform assessment development
Early number	Aubrey <i>et al.</i> , 2006 Jordan <i>et al.</i> , 2009 Nguyen <i>et al.</i> , 2016
Early calculation	Aubrey <i>et al.</i> , 2006 Casey <i>et al.</i> , 2017 Jordan <i>et al.</i> , 2009 Mulligan and Mitchelmore, 2009
Mathematical language	Aubrey <i>et al.</i> , 2006 Duncan <i>et al.</i> , 2007
Early understanding of pattern	Mulligan and Mitchelmore, 2009

Table 10: Supporting research evidence for mathematics content domain

Content domain: Literacy, Communication and Language

The development of the LCL tasks of the RBA was underpinned by research showing that language development is central to a child’s ability to access the curriculum and develop literacy skills (for example, Bowman *et al.*, 2000; Cooper *et al.*, 2002). In developing the assessment tasks, consideration was given to current theories of language development, for example evidence in relation to vocabulary (Stadthagen-Gonzalez and Davis, 2006; Cooper *et al.*, 2002), and the important impact of phonological awareness on the development of children’s reading skills (Scanlon and Vellutino, 1996).

The table below provides a summary of the LCL content domains, together with the research that underpinned their development.

Content domain	Supporting research evidence used to inform assessment development
Early vocabulary	Cooper <i>et al.</i> , 2002 Dogan <i>et al.</i> , 2015 McGill-Franzen, 2010 Muter <i>et al.</i> , 2004 Stadthagen-Gonzalez and Davis, 2006
Phonological awareness	Dogan <i>et al.</i> , 2015 Muter <i>et al.</i> , 2004 Savage and Carless, 2008 Scanlon and Vellutino, 1996 Walker <i>et al.</i> , 2015
Early comprehension	Oakhill <i>et al.</i> , 2003

Table 11: Supporting research evidence for LCL content domain

1.3 Does the cognitive domain provide an accurate reflection of the scope of learning at the start of reception?

Cognitive development during the early years encompasses a wide range of cognitive skills. Young children are generally active and engaged pupils who learn through exploration and possess a natural curiosity. The RBA seeks to reflect the types of learning that children will have experienced at the start of reception.

The [Assessment framework – Reception Baseline Assessment](#) (page 7) provides information about the cognitive domain, explaining that *the reception baseline is a task-based assessment designed to maximise the active interaction between the pupil and the resources and that early cognitive development is multifaceted and the reception baseline requires a range of cognitive processes to be used in responding to the variety of tasks included.*

The RBA allows pupils to respond to the assessment in a variety of ways such that they *may need to demonstrate linguistic skills such as blending sounds, or mathematical skills such as number sense. Some cognitive processes, such as memory or attention, are deployed in both the mathematics and LCL tasks.*

During development, the tasks were scrutinised by a variety of panels, including, a practitioner panel ($n=11$) and an expert panel ($n=7$) (both convened by NFER). A stakeholder group ($n=18$) (convened by DfE) also reviewed and provided feedback on the assessment items, although this was not the main focus of the group. From all panels, feedback indicated that the tasks were appropriate for children of this age group and reflective of children’s scope of learning at this stage of their development. This took into consideration that, as a standardised assessment, all assessment activities had to be designed in such a way that they could be consistently administered within and across

schools. Therefore, the levels of personalisation of activities that are accepted practice in the Early Years would not be appropriate for the reception baseline assessment. Practitioner feedback from trialling also indicated that the tasks reflect the scope of learning (see section 1.2).

1.4 Have tasks been rigorously reviewed for their validity by a range of appropriate stakeholders? To what extent has feedback led to refinements of tasks?

The assessment development process is designed to ensure regular review and validation by a range of stakeholders throughout the development of an assessment. For the RBA, these stages were:

- a) Item writing – NFER undertook internal review of the items written, before trialling them in a small number of schools. These initial items were also reviewed by STA's test development researchers and external curriculum advisors. Suggestions for improvements to items were agreed before they were further developed in preparation for the trial.
- b) Expert review – items were reviewed by a wide range of stakeholders to confirm their suitability. An expert panel ($n=7$), which included assessment, special educational needs and disabilities (SEND) and early years experts, and a teacher panel ($n=11$) were convened. Other stakeholders ($n>38$) including teachers, subject and early years experts (including the DfE early years team), SEND experts, inclusion experts, assessment experts and LA staff also examined and commented on the materials at various stages of development. Feedback from all reviewers was collated by NFER's assessment researchers and decisions were taken on improvements to the items in conjunction with STA staff.
- c) Trialling – feedback was sought from practitioners who trialled the assessment ($n=313$) and this informed the next stage of development. Practitioner comments were reviewed alongside the data from the assessment. Where it was deemed necessary and appropriate, changes were agreed with STA as part of the assessment finalisation process in preparation for the pilot.
- d) Pilot – during the first half of the Autumn term, 7,046 schools completed assessments for all pupils in their reception cohort. Assessment data was gathered for 339,477 pupils. Practitioners were invited to complete surveys in order to provide feedback on the assessment and the training. Feedback from practitioners in relation to the assessment ($n=2,507$) confirmed the suitability of the tasks. Practitioner comments were reviewed alongside observational data and statistical data from the assessment. Where it was deemed necessary and appropriate, some minor changes to the assessment tasks were agreed with STA as part of the assessment finalisation process.

1.5 Have tasks and responses from trialling been suitably interrogated to ensure only the desired construct is being assessed (and that construct irrelevant variance is minimised)?

Following the trial, all tasks, items, pupil responses and practitioner feedback were

reviewed. In particular, there were two tasks where it was necessary to review all the answers that children gave to determine which answers were creditworthy and to refine the marking criteria and exemplar answers. Alongside these qualitative reviews, item statistics were also analysed. The discrimination statistics, in particular, helped to determine whether or not items had functioned as expected. Items with low discrimination values were examined to ensure that the item was not subject to construct irrelevant variance. Analysis of differential item functioning (DIF) was also carried out to check for construct irrelevant bias (see section 2.4). Outcomes from all of these reviews helped to inform the decision making process as to whether or not items were suitable for selection. Additional information is also provided in section 2.5.

Evidence from the pilot has also been reviewed and changes were made to the assessment to minimise assessment length whilst maintaining assessment quality.

1.6 Does the final assessment provide sufficient coverage of the content domain (whilst meeting the requirements within the assessment framework)? Is a range of tasks included that are appropriate to classroom practice?

The content domains outlined in the [Assessment framework](#) constitute the entirety of the assessment. As part of the item selection process, NFER and STA looked at the balance between the assessment components (LCL and mathematics) and also the balance across and between the content domains.

These can be summarised as follows:

Mathematics component	Approximate proportion of all assessment items*
Early number	50%
Early calculation (early addition/subtraction)	30%
Mathematical language	10%
Early understanding of pattern	10%

Table 12: Approximate proportion of assessment items by mathematics component

* Note that due to routing, not all children will see all items. This table refers to all items within the assessment component.

LCL component	Approximate proportion of all assessment items*
Early vocabulary	40%
Phonological awareness	45%
Early comprehension	15%

Table 13: Approximate proportion of assessment items by LCL component

* Note that due to routing, not all children will see all items. This table refers to all items within the assessment component.

Proportions of the different content domains were also compared with the proportions outlined in the [Key stage 1 and 2 assessment frameworks](#).

Another consideration when selecting the tasks was the balance of the demands on practitioner and child time, whilst collecting sufficient information to inform the measure. Therefore item information and timing data were considered in tandem as part of the item selection process.

A range of tasks were selected that reflect aspects of classroom practice. The child-centred assessments are carried out one-to-one with a practitioner. The tasks are practical and there is no written component to the assessment. Careful consideration was given to the balance of both concrete and abstract tasks in order to ensure that the final assessment fully aligns with the [Assessment framework](#).

Claim 2

Assessment results provide a fair and accurate measure of pupil performance.

Research questions for claim 2

2.1 How has item level data been used in assessment construction to ensure only items that are functioning well are included in the assessment?

Item performance statistics (classical and item response theory) from the trial were reviewed in meetings between researchers and psychometricians to initially determine which items were performing sufficiently well to potentially be included in the live assessment. Evidence related to individual items was reported to STA. Due to the assessment being confidential, this information is not available publicly.

An item selection meeting then took place involving NFER and members of STA's assessment development division. The purpose of the meeting was to review all available evidence and make decisions on the most appropriate next stage for each task and each item with each task.

The first stage was to review the evidence for each task as a whole and to determine whether the task was viable. For each task, consideration was given as to whether or not the task worked in principle and, if so, whether the items within a task were suitable for use in the final assessment. Consideration was also given to the associated resources or stimulus materials to determine whether any amendments were needed.

For each item within a task, the following evidence was reviewed:

- Classical Test Theory (CTT) and Item Response Theory (IRT) analysis of the performance of items in order to determine characteristics such as their difficulty and discrimination
- Differential Item Functioning (DIF) analysis by gender and EAL status
- Analysis of individual interactions (distractor analysis)
- Reviews of children's responses to open-ended items to see how children were interacting with those questions
- Relationship with other items within the trial in the form of factor analysis and correlation between residuals from the IRT model (Q3 statistic)
- Distribution of time taken to administer the group of items within the same screen of the assessment (in the form of deciles)
- Feedback from practitioners involved in the trial and feedback received from panel meetings

After review, the following outcomes were available for each item:

- Item is available for selection in the final assessment unamended since there is

- sufficient evidence that the item is performing as intended
- Item is available for selection in the final assessment with amendments since, although there is some evidence that the item is not performing as intended, the issue has been identified and corrected
 - Remove the item as major issues have been identified that cannot be corrected

Any item that was determined to be available for inclusion in the final assessment has therefore demonstrated that it assesses the appropriate construct and meets the required item functioning criteria.

Additionally, since trial analysis had revealed that there were high correlations between some items, this information was used to avoid selecting items which were too strongly related and could therefore be considered to be assessing the same element of the construct. The inclusion of this information helped to ensure the coverage of the content domain could be as broad and balanced as possible.

Item information functions resulting from the IRT analysis were then used to produce a theoretically optimal assessment in terms of psychometric functioning within the constraint of 70% of the trial sample being able to complete the assessment within around 20 minutes. This selection was compared at item level to the initial selection in order to provide challenge for the decisions made and to ensure they were robust.

Further evidence on item and whole assessment performance was collected during the pilot and analysis confirmed that the assessment items functioned appropriately. For the amended assessment, the timing data suggests that the majority of administrations (92%) will be completed within 20 minutes.

2.2 How has qualitative data been used in assessment construction to ensure only items that are effectively measuring the desired construct are included in the assessment?

A range of qualitative data sources were taken into account as part of the construction of the assessment. NFER convened both practitioner and expert panels. Items were interrogated by these groups at three stages in the development process: firstly during the initial development of the items, then after the informal trial in schools at which point items were amended. Finally, the groups were presented with the data and suggested revisions after the September – October 2018 trial. At all points, both groups made valuable contributions that informed the decision-making process for the final assessment construction. Additionally, the practitioner panel met in November 2019 to look at findings from the first half term of the pilot.

NFER researchers also sought feedback directly from schools (both practitioners and children) on the assessment at the initial stage of development (informal trialling), during the September – October 2018 trial, and during the pilot year. Researchers undertook the assessment with children and also observed practitioners carrying out the assessment.

The feedback from children and practitioners was used to inform the final assessment construction.

Feedback from the practitioner panel meeting and from the other qualitative data sources, including practitioner feedback (n=45), contributed to further refinement of the assessment and associated guidance as part of the finalisation process.

In addition, STA provided feedback on the assessment at all stages of the development process and was actively engaged with reviewing all evidence during the construction of the final assessment.

As described in 1.2 above, considerably more items were trialled than were needed for the final RBA. As a result of the reviewing described above, about half of the LCL items and about a third of mathematics items were removed from the final selection.

2.3 Are a range of valid and age appropriate items included that allow children of the full ability range to demonstrate their capabilities?

During the September – October 2018 trial, practitioners were asked to rate the appropriateness of various aspects of the assessment on a scale ranging from ‘very poor’ to ‘very good’.

Practitioners rated the suitability of practical resources highly, with large numbers rating this as ‘satisfactory’ or above (95% for those trialling mathematics tasks and 93% for those trialling LCL tasks) and made comments such as *‘resources were appropriate for age group’*. More than four-fifths (82%) of practitioners trialling the mathematics tasks rated them as ‘satisfactory’ or above, making comments such as *‘The Maths tasks were very good, very similar to the type of activities I do with new reception children every year to assess where they are. The children enjoyed them too’*. Nearly two-thirds (65%) of practitioners trialling the LCL tasks considered them to be at least ‘satisfactory’.

In response to feedback from the trial, the provision of resources was reviewed to ensure that they would be as manageable for the practitioner and as supportive for the child as possible and some changes in provision were made.

The IRT test information function was reviewed for the selected items against the ability distribution for the trial pupils in order to ensure that the assessment would measure sufficiently well across the full ability range. Expected item facilities and total score distribution were also produced in order to check that there were low numbers of pupils achieving 0 or full marks, thereby ensuring the assessment had sufficient easy items for lower ability pupils to access and sufficient difficult items to avoid a ceiling effect, given the intended purpose of the assessment as a baseline for a progress measure.

Further evidence on item and whole assessment performance, in particular the score distribution for the assessment as a whole, was collected during the first half term of the pilot. Whilst it is not appropriate to include the score distribution here as the purpose of

the RBA is to measure a school's impact on pupil progress and not to make inferences about individual pupils, the percentage of pupils scoring zero and full marks shall be made available. Analysis of the pilot (prior to assessment changes) confirmed that there was a good spread of pupils across the score range, with less than 0.7% of pupils scoring no marks and less than 0.1% achieving full marks. This provides evidence that there is not a ceiling effect on the assessment and that it can discriminate well between pupils across the ability range.

The internal reliability of the assessment was measured using Cronbach's Alpha. An Alpha coefficient of 0.7 or above is generally considered sufficient for an assessment to be considered suitable to use for drawing inferences about groups. Following changes to the assessment (see 1.2), the Cronbach's Alpha for the whole assessment is predicted to be 0.91, demonstrating a high degree of internal consistency reliability.

2.4 What evidence has been used (qualitative and quantitative) to ensure the assessment does not disproportionately advantage or disadvantage any sub-groups?

The relationships between pupil gender and age and performance on the assessment were analysed. Girls significantly outperformed boys on the overall assessment as well as on the individual components. A clear trend between age (in months) and performance can also be observed in the data, with older pupils outperforming the younger members of the cohort. This is true for the overall assessment as well as for the individual components. However, differences in performance of sub-groups is not necessarily a sign of bias in an assessment and may just reflect differences in attainment. In addition, as the RBA is the starting point for a cohort level progress measure and will be used to place children into comparison groups with others with a similar prior attainment, such differences are not a cause for concern. They would only be a concern if the assessment was designed to be an individual or absolute measure of attainment at this age.

To investigate bias, different statistical and judgemental processes are used.

As part of the 2018 trial, differential item functioning (DIF) analysis was carried out by gender and EAL in order to highlight items where those groups performed differently relative to their overall ability. The items flagged as exhibiting DIF were then reviewed to ensure that the difference in performance was not due to construct irrelevant bias.

In respect of gender, the LCL component did not contain any items which exhibited differential item functioning in the trial. Whilst the mathematics component did contain some items which exhibited DIF in relation to gender, there was a balance across boys/girls. In terms of items showing differential item functioning by EAL/non-EAL, there was a balance across the two groups and across LCL and mathematics.

DIF analysis by gender and EAL was also carried out using data from the first half term of the pilot. This revealed that both components had one item which exhibited differential

item functioning in favour of boys. This should be interpreted in the context that girls performed better on average than boys on both components overall. These two items were both subject to routing and in both cases a larger proportion of girls was presented with the item but, of the pupils attempting the item, a larger proportion of boys answered correctly. There was no evidence to suggest these differences could have been due to any construct irrelevant bias and are therefore not considered a threat to the validity of the assessment.

The EAL DIF analysis indicated ten LCL items (of which 7 were in favour of pupils with English as a first language) and seven mathematics items exhibited DIF (all in favour of pupils with EAL). Eight of the LCL items were noted as previously displaying DIF in the trial, as were four of the mathematics items, and all were deemed not to be suffering from construct irrelevant bias. Additionally, pupils with EAL are expected to perform less well on language-based items than numeracy-based items (as is the case in the 2019 pilot) and this can be seen when mean score on component is broken down by EAL status. It is therefore considered that there is no evidence that the validity of the assessment is undermined by the results of this analysis.

As part of the development process, and separate to the practitioner and expert panels, the materials were reviewed by both a cultural reviewer and a SEND reviewer.

The SEND reviewer commented that the assessment '*shows an excellent regard for the barriers that SEND children may face*'. The appropriacy of the assessment materials was also remarked on as the reviewer considered that SEND pupils should be able to access some if not all of the assessment. Furthermore it was felt that the guidance documentation sets high expectations for pupils with SEND, contrary to the general tendency to assume that pupils with SEND will perform poorly.

The cultural reviewer considered that '*All the materials are acceptable from a cultural point of view*' and are '*... unproblematic across a wide spectrum of religious and ethnic communities*'. The inclusive nature of the images was also remarked upon since they '*include variations of skin tone without exaggerating physical differences.*'

Although it was recognised that children with EAL may have additional difficulties with the assessment, it was not felt that any of the assessment content needed either to be removed or simplified. The removal of items would mean that it would not be possible to ensure coverage of all content domains for all children. In addition, at all stages of the process, question wording was reviewed and simplified as far as possible.

The assessment was also reviewed by SEND specialists who have provided guidance and developed modified materials for the assessment which are suitable for children with hearing and/or visual impairments. These are available to schools upon request.

Qualitative data, including information from observations and feedback from practitioners, was used to develop and refine guidance provided to support access for children with additional needs (e.g. additional SEND advice and EAL guidance).

2.5 Have pupil responses been interrogated to ensure pupils are engaging with the tasks as intended?

During the trial, practitioners ($n=313$) were asked to consider children's understanding of the tasks. The majority of practitioners rated children's understanding of the tasks as 'satisfactory' or above (94% for mathematics, 88% for LCL), making comments such as:

- *The children understood most activities. They enjoyed playing the games.*
- *The practical resources were effective in supporting the questions.*
- *Tasks were generally pitched appropriately and gave useful insights.*

For two tasks where there was opportunity for variation in responses, pupil answers were collated and coded to ensure that the associated guidance would clearly support the range of answers that might be given.

The review process also helped to identify any tasks which demonstrated that children had not engaged with the task as intended. This resulted in one task being removed from the final selection and ensured that scoring guidance was amended, where necessary, to reflect the types of responses being given.

Feedback from practitioners ($n=2,507$) during the first half term of the pilot confirmed that the majority of them (89%) considered that the children's understanding of the tasks was 'satisfactory' or better. Qualitative comments from practitioners suggested that the tasks were a useful assessment tool.

As part of the observational and quality monitoring visits⁴ during the pilot, practitioners were asked to identify if there were any questions which gave rise to responses that they considered to be valid but which were not covered in the guidance. Two tasks were identified that would benefit from further guidance/exemplification in the scoring guidance and amendments were agreed with STA for the live assessment.

Feedback from observational visits and the practitioner panel indicated that pupils demonstrated high levels of engagement with the tasks and that the items were interpreted as intended.

⁴ There were also visits carried out by trained quality monitors ($n=98$). The role of the quality monitors, all of whom were experienced early years teachers / headteachers, was to visit a representative range of schools and carry out systematic observations and interviews with practitioners and headteachers. This provided data on practitioner experience, attitudes and behaviours and system level effectiveness. These visits also determined how far the assessment was carried out as intended and identified any areas of uncertainty or ambiguity with regard to administration of the RBA.

2.6 Is the rationale for what is creditworthy robust and valid? Can this rationale be applied unambiguously?

The item design ensures that practitioners can easily make binary decisions on answers given by the children. Questions are accessible and scoring rubrics indicate the required responses. The majority of tasks have only one correct answer. Where different response strategies are permissible, scoring 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.

At all stages of the assessment development process, creditworthy answers were scrutinised and validated by NFER researchers, STA, and the practitioner and expert panels.

For one question, answers were collected at trial to construct the guidance about creditworthy responses and to provide exemplar answers. In order to support practitioners in the decision-making process and support consistency of scoring, this task is also exemplified in the training videos.

2.7 Are tasks trialled to ensure that all responses showing an appropriate level of understanding are credited and that no responses demonstrating misconceptions or too low a level of understanding are credited? Does the guidance provide appropriate detail and information for practitioners to make decisions confidently and reliably?

The assessment requires practitioners to make a binary decision against each item. During the trial, practitioners were asked if there were any tasks where it was difficult to decide whether to select 'yes' or 'no'. More than four-fifths of practitioners (87% of those trialling mathematics tasks and 86% of those trialling LCL tasks) considered it was not difficult to make the decision. Following the trial, some changes were made to two tasks in the online scoring system to help support practitioners in recording their yes/no decision-making. This was further supported by additional guidance and training materials. Assessment survey feedback from the first term of the pilot indicated that the majority of practitioners (83%) considered that the guidance for awarding 'yes' or 'no' for each item was 'good' or 'very good'.

The vast majority of questions have a defined set of creditworthy answers and these are supported by exemplification of 'do not accept' responses.

Tasks were reviewed throughout the development process and practitioners had the opportunity to identify any parts of the yes/no decision-making guidance that was considered to be ambiguous or unclear. Very few comments were made but consideration was given to those received and amendments were made accordingly.

During the trial, practitioners were asked to record children's answers for some questions where there was more than one correct response. These responses were then collated and used to inform the development of the acceptable points in the guidance.

On-screen item types have been designed to ensure that practitioners are able to record 'yes' or 'no' consistently and accurately. For example, in some questions, depending on the practitioner's choices, other options are pre-populated to prevent mis-scoring the data. The functionality of these item types was discussed with both expert and practitioner panels and with STA. Further support is given to practitioners through the training videos.

Furthermore, in 2015 a study was carried out based on the NFER's Reception Baseline Assessment (on which the RBA is based). A sample of 150 children were assessed twice in the same week by different practitioners to provide some evidence of test/re-test reliability and marker/assessor reliability. Since these two aspects are confounded it is not possible to separate out the two effects. The correlation between the two sets of scores achieved in the study was 0.96, demonstrating that the outcomes from the assessment were very stable under repeated administration within a short timeframe (a correlation of 0.7 is the accepted threshold for test-retest reliability). At item level, the agreement between practitioners was also high, with an average agreement of 87%. However, since the response assigned by the practitioner was binary, this figure will be affected by some degree of agreement by chance. Despite the limitations of the study, the evidence indicated that an assessment very similar in nature to the new RBA had a good degree of reliability under test/re-test conditions with a different practitioner.

For one mathematics item, observations of assessments carried out in the first half term of the pilot revealed a few variations in the possible responses that could arise. Discussions with the practitioner panel confirmed which answers should be creditworthy according to early years practice, and these were added to the scoring guidance.

Practitioner feedback ($n=2011$) on the pilot indicated that the majority of practitioners (94%) considered that the administration guide was 'quite useful' or better, with 88% indicating that it contained 'the right amount of information'. Feedback from quality monitor interviews ($n=141$) confirmed that 91% of practitioners felt confident carrying out the assessments.

2.8 Are practitioners making use of the task guidance to support decisions, as intended?

Evidence from observational and quality monitoring visits during the first half term of the pilot indicated that the majority of practitioners did make use of task guidance to support decisions, as intended. Feedback from these visits did prompt further refinement of some of the general guidance with regard to approaches to tasks. The feedback also led to minor wording changes to some items to ensure clarity and to support consistency in delivery.

Claim 3

Pupil performance is comparable within and across schools.

Research questions for claim 3

3.1 Is potential bias to particular sub-groups managed and addressed when constructing the assessment?

During development, the tasks have been reviewed by a number of external experts, as well as by the expert panel, to ensure that they would not unfairly advantage or disadvantage particular groups of children. The review included considering children with SEND and materials were examined from a cultural perspective to ensure that there was no bias. Any comments made by reviewers fed into the development process for the items.

As discussed in section 2.4, differential item functioning analysis was carried out by gender and EAL. Information from this analysis was considered during item review and selection, to ensure that any performance differences were not due to construct irrelevant bias and that there was balance across the different sub-groups.

3.2 Are systems in place to ensure the security of assessment materials during development and administration?

The RBA was developed and trialled in line with the usual confidential assessment development requirements.

When schools signed up for the pilot, they signed a confidentiality agreement committing to the continued safe-keeping of all resources and preventing any member of staff from discussing the content of the assessment in any forum, including on social media.

Materials are despatched from the secure warehouse facility and sent on an overnight dedicated courier, with GPS tracking, to a named contact at each school.

Schools are required to store the RBA materials securely, in line with requirements for other confidential assessments.

Schools are provided with a username for the Baseline e-portal (BeP), which can only be accessed with a password.

Going forwards, headteachers will be required to sign a headteacher declaration form stating that they will be responsible for secure storage of the materials and for maintaining the confidential status of all aspects of the RBA.

STA monitors media stories and social media and ensures appropriate action is taken to address security concerns.

3.3 Is administration guidance available, understood and implemented consistently across schools?

Practitioners are supported in administering the assessment through the use of a printed administration guide, preview assessments on the BeP and a series of training videos which demonstrate some tasks and how to use the online system.

The vast majority of practitioners (in excess of 90%) who participated in the trial considered that the administration guide provided all the information required to confidently administer the assessment. This finding was confirmed during the pilot.

Practitioner feedback ($n=2,011$) from the pilot indicated that the majority of them found the training materials useful (92%) and considered that an appropriate amount of information was provided in the administration guide (88%) and within the training videos (88%). The training materials were accessed by the majority of schools (95%) who logged onto the system during the first half term of the pilot. Feedback from the quality monitoring visits confirms that the majority of practitioners followed the general guidance set out in the administration guide and were suitably prepared to carry out the assessments.

To further improve the guidance, the following recommendations will be implemented:

1. Provide additional guidance in the administration guide and/or item level guidance specifying how resources should be set up and managed within or between items.
2. Exemplify ways of using the training materials to best prepare for the RBA, including how best to adapt the assessment 'in house' for children with additional needs and the roles and responsibilities of the headteacher.
3. Provide additional guidance to further support the consistency of delivery:
 - include further guidance about which elements of the question can be repeated, if necessary
 - provide additional guidance about how and when a child can be prompted to answer
 - provide additional guidance detailing how to clarify a child's answer where this may be unclear or ambiguous
 - increase the emphasis in the administration guide relating to the need to follow the specified wording in the question
 - provide additional guidance about the use of praise and encouragement during tasks
4. Ensure clear messaging about the six-week window for carrying out the assessment being from the time that an individual child starts school.
5. Provide clear guidance about use of the 'disapply', 'pause', 'discontinue' and 'delay' functions.
6. Ensure schools are aware of the minimum operating requirements.

7. Provide clear guidance about the need to create accounts for every user in order to avoid users sharing a single account.
8. Make common transfer file (CTF) the default method of uploading pupil data, alongside clear guidance on how to use this method.
9. Alert practitioners to any dates of birth outside a specific range in order to highlight potential data upload errors. Add a requirement for the headteacher to declare that all dates of birth on the system are accurate to the best of their knowledge.

3.4 Are the available access arrangements appropriate?

The assessment has been designed to be accessible to the vast majority of children on entry to reception. However, the assessment has also been subject to review by SEND experts and has had a specific review in relation to children with visual and hearing impairments.

Guidance and modified resources are available to support practitioners who are assessing children with hearing and/or visual impairment.

A full range of access arrangements will be made available for schools to use, in line with other national assessments. Full details of the permitted access arrangements will be made available in the RBA assessment and reporting arrangements document that will be published in April 2020.

Claim 4

The meaning of total scores is clear to those responsible for creating the progress measures.

Research questions for claim 4

4.1 Is appropriate guidance available to ensure the DfE data team understand the reported scores?

The data collected as part of the RBA and sent to the DfE data teams includes raw scores and some pupil information (as specified in the privacy notice). This data is under field headings that are simple to understand and that have been agreed with the DfE data teams. The data specification adds further information about this data, to ensure that the DfE data team understand what has been collected. There are no reported scores produced for the RBA by the STA or the DfE. Instead, the raw scores are stored in the National Pupil Database for use seven years later in producing the progress measure.

4.2 Are the unintended consequences of the assessment known, and have threats to validity been investigated and mitigated?

Possible unintended consequences of the assessment have been considered as part of the formulation of this new policy and, in particular, the unintended consequences that might arise from what is done with the raw scores and with whom this data is shared. Two main risks were identified here: the potential for streaming or labelling of the children based on their scores, and the potential for retrospective judgement of early years provision.

To mitigate these risks, DfE will not share raw scores with schools, teachers or parents. This data will be stored in the National Pupil Database, and will only be used to form the progress measure at the end of key stage 2. This will help to prevent scores being used as a grouping mechanism, and it will mean that there is a reduced risk of any early years settings being assessed based on a school's RBA total scores. This will also help to reinforce the important message that no preparation is necessary ahead of the assessment, and that neither schools nor parents need to do any practice beforehand.

This assessment is not designed to be used for any purpose other than the progress measure. It is intended to assess a child's early literacy, communication and language and mathematics to form the start of the cohort-level progress measure. The raw scores are not intended to be used for any diagnostic purposes or outcome measures. This has been communicated clearly in the information published by DfE about the assessment.

References

- Aubrey, C., Godfrey, R. and Dahl, S. (2006). 'Early mathematics development and later achievement: further evidence', *Mathematics Education Research Journal*, **18**, 1, 27–46 [online]. DOI 10.1007/BF03217428.
- Bowman, B., Donovan, M.S. and Burns, M.S. (Eds) (2000). *Eager to Learn: Educating Our Preschoolers*. Washington DC: The National Academy Press.
- Casey, B.M., McPherran Lombardi, C., Pollock, A., Fineman, B. and Pezaris, E. (2017). 'Girls' spatial skills and arithmetic strategies in first grade as predictors of fifth-grade analytical math reasoning', *Journal of Cognition and Development*, **18**, 5, 530–555 [online]. DOI 10.1080/15248372.2017.1363044.
- Cooper, D.H., Roth, F.P., Speece, D.L. and Schatschneider, C. (2002). 'The contribution of oral language skills to the development of phonological awareness', *Applied Psycholinguistics*, **23**, 3, 399–416 [online]. DOI 10.1017/S0142716402003053.
- Dogan, E., Ogut, B. and Kim, Y. Y. (2015). 'Early childhood reading skills and proficiency in NAEP eighth-grade reading assessment', *Applied Measurement in Education*, **28**, 3, 187–201 [online]. DOI 10.1080/08957347.2015.1042157.
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. and Japel, C. (2007). 'School readiness and later achievement', *Developmental Psychology*, **43**, 6, 1428–1446 [online]. DOI 10.1037/0012-1649.43.6.1428.
- Early Education (2012). *Development Matters in the Early Years Foundation Stage (EYFS)* [online]. Available: [Development Matters in the Early Years Foundation Stage: Non-statutory guidance material \(PDF\)](#) [6 September, 2019].
- Jordan, N.C., Kaplan, D., Ramineni, C. and Locuniak, M.N. (2009). 'Early math matters: kindergarten number competence and later mathematics outcomes', *Developmental Psychology*, **45**, 3, 850–867 [online]. DOI 10.1037/a0014939.
- McGill-Franzen, A. (2010). 'Guest editor's introduction', *Educational Researcher*, **39**, 4, 275–278 [online]. DOI 10.3102/0013189X10370619.
- Mulligan, J. and Mitchelmore, M. (2009). 'Awareness of pattern and structure in early mathematical development', *Mathematics Education Research Journal*, **21**, 2, 33–49 [online]. DOI 10.1007/bf03217544.
- Muter, V., Hulme, C., Snowling, M.J. and Stevenson, J. (2004). 'Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study', *Developmental Psychology*, **40**, 5, 665–681 [online]. DOI 10.1037/0012-1649.40.5.665.
- Nguyen, T., Watts, T.W., Duncan, G.J., Clements, D.H., Sarama, J.S., Wolfe, C. and Spitler, M.E. (2016). 'Which preschool mathematics competencies are most predictive of fifth grade achievement?', *Early Childhood Research Quarterly*, **36**, 550-560 [online]. DOI 10.1016/j.ecresq.2016.02.003.
- Oakhill, J.V., Cain, K. and Bryant, P.E. (2003). 'The dissociation of word reading and text comprehension: evidence from component skills', *Language and Cognitive Processes*, **18**, 4, 443–468 [online]. DOI 10.1080/01690960344000008.

Savage, R. and Carless, S. (2008). 'The impact of early reading interventions delivered by classroom assistants on attainment at the end of year 2', *British Educational Research Journal*, **34**, 3, 363–385 [online]. DOI 10.1080/01411920701609315.

Scanlon, D.M. and Vellutino, F.R. (1996). 'Prerequisite skills, early instruction, and success in first-grade reading: selected results from a longitudinal study', *Mental Retardation and Developmental Disabilities Research Reviews*, **2**, 1, 54–63 [online]. DOI 10.1002/(SICI)1098-2779(1996)2:1<54:AID-MRDD9>3.0.CO;2-X.

Stadthagen-Gonzalez, H. and Davis, C. J. (2006). '*The Bristol norms for age of acquisition, imageability, and familiarity*', *Behavior Research Methods*, **38**, 4, 598-605 [online]. DOI 10.3758/BF03193891.

Walker, M., Sainsbury, M., Worth, J., Bamforth, H. and Betts, H. (2015). *Phonics Screening Check Evaluation: Final Report* [online]. Available: [Phonics screening Check Evaluation: Final report \(PDF\)](#) [6 September 2019].



© Crown copyright 2020

This publication (not including logos) is licensed under the terms of the Open Government Licence v3.0 except where otherwise stated. Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned.

To view this licence:

visit www.nationalarchives.gov.uk/doc/open-government-licence/version/3

email psi@nationalarchives.gsi.gov.uk

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

About this publication:

enquiries www.education.gov.uk/contactus

download www.gov.uk/government/publications

Reference: 978-1-78957-516-3

STA/20/8533/e



Follow us on Twitter:
[@educationgovuk](https://twitter.com/educationgovuk)



Like us on Facebook:
facebook.com/educationgovuk